# Financial Time Series Forecasting Using Improved Wavelet Neural Network

Master's Thesis

Chong Tan
20034244

Supervisor
Prof. Christian Nørgaard Storm Pedersen

May 31, 2009

# Abstract

In this thesis, we propose an improved exchange rate forecasting model based on neural network, stationary wavelet transform and statistical time series analysis techniques. We compare the new model's performance with pure neural network forecasting model, wavelet/wavelet-packet-denoising-based forecasting models and the methods used in nine previous studies. The experimental results demonstrate that the proposed method substantially outperforms existing approaches.

# Acknowledgements

I would like to thank to my supervisor, Prof. Christian Nørgaard Storm Pedersen, for his guidance and support throughout my thesis work, and I want to express my gratitude to my former supervisor, Prof. Brian H. Mayoh, for his help, advice and encouragement over the past several years.

*Chong Tan,*
*Århus, May 31, 2009.*

# Contents


**Abstract** **3**

**Acknowledgements** **5**

**1 Introduction** **5**
1.1 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

**2 Neural Network** **11**
2.1 Processing Unit . . . . . . . . . . . . . . . . . . . . . . . . . . 11
2.2 Neural Network Topologies . . . . . . . . . . . . . . . . . . . . . 13
2.3 Learning and Generalization . . . . . . . . . . . . . . . . . . . . 13

**3 Wavelet Transform** **17**
3.1 Why Wavelet Transform ? . . . . . . . . . . . . . . . . . . . . . 17
3.2 Continuous Wavelet Transform (CWT) . . . . . . . . . . . . . . . 17
3.3 Discrete Wavelet Transform (DWT) . . . . . . . . . . . . . . . . 18
3.4 Stationary Wavelet Transform (SWT) . . . . . . . . . . . . . . . 20
3.5 Wavelet Packet Transform . . . . . . . . . . . . . . . . . . . . . 20
3.6 Wavelet and Wavelet Packet Denoising . . . . . . . . . . . . . . . 20

**4 Forecast Model** **25**
4.1 Overview of the Design . . . . . . . . . . . . . . . . . . . . . . . 25
4.2 Network Structure Design . . . . . . . . . . . . . . . . . . . . . . 28
4.3 Wavelet Features . . . . . . . . . . . . . . . . . . . . . . . . . . 28
4.4 Statistical Features . . . . . . . . . . . . . . . . . . . . . . . . . 30
4.4.1 Mean . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
4.4.2 Mean Absolute Deviation . . . . . . . . . . . . . . . . . . 30
4.4.3 Variance . . . . . . . . . . . . . . . . . . . . . . . . . . 31
4.4.4 Skewness . . . . . . . . . . . . . . . . . . . . . . . . . . 31
4.4.5 Kurtosis . . . . . . . . . . . . . . . . . . . . . . . . . . 32
4.4.6 Turning Points . . . . . . . . . . . . . . . . . . . . . . . 32
4.4.7 Shannon Entropy . . . . . . . . . . . . . . . . . . . . . . 32
4.4.8 Historical Volatility . . . . . . . . . . . . . . . . . . . . . 33

**5 Experimental Results and Discussions** **35**
5.1 Mackey-Glass Time Series . . . . . . . . . . . . . . . . . . . . . 35
5.2 Exchange Rates Forecasting . . . . . . . . . . . . . . . . . . . . 37

5.3 Comparison With Other Research . . . . . . . . . . . . . . . . . . 51

**6 Conclusion and Future Work** **81**

**A Appendix** **83**
A.1 Statistical Performance Measures . . . . . . . . . . . . . . . . . . . 83
A.2 Middle-term and Long-term Forecasting Results . . . . . . . . . . 84

**Bibliography** **97**

# List of Figures


2.1 Artificial Neuron Model . . . . . . . . . . . . . . . . . . . . . . . . 11
2.2 Activation Functions . . . . . . . . . . . . . . . . . . . . . . . . . . 12
2.3 Multilayer Feedforward Neural Network . . . . . . . . . . . . . . 13
2.4 Overfitting in Supervised Learning . . . . . . . . . . . . . . . . . . 14

3.1 Filter Bank Scheme for DWT . . . . . . . . . . . . . . . . . . . . . 19
3.2 Wavelet Decomposition Tree . . . . . . . . . . . . . . . . . . . . . . 19
3.3 Filter Bank Scheme for SWT . . . . . . . . . . . . . . . . . . . . . 20
3.4 Wavelet Packet Decomposition Tree . . . . . . . . . . . . . . . . . 21
3.5 Level 4 Stationary Wavelet Decomposition using Haar Wavelet . 22
3.6 Level 4 Discrete Wavelet Analysis using Haar Wavelet . . . . . . 23

4.1 Pure Neural Network Forecasting Model . . . . . . . . . . . . . . 25
4.2 Wavelet/Wavelet Packet Denoising-based Forecasting Model . . . 25
4.3 Proposed Model: Training . . . . . . . . . . . . . . . . . . . . . . . 26
4.4 Proposed Model: Forecasting . . . . . . . . . . . . . . . . . . . . . 26
4.5 An Illustration of Different Forecasting Modes . . . . . . . . . . . 27
4.6 The Haar Wavelet* . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
4.7 The Db40 Wavelet . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
4.8 A Comparison of the Wavelet and Wavelet-Packet Denoising Techniques (Db40 wavelet function is used) . . . . . . . . . . . . 31

5.1 Mackey-Glass Time Series . . . . . . . . . . . . . . . . . . . . . . . 36
5.2 Forecasting result of Mackey-Glass Time Series (1-step-ahead) . . 39
5.3 Forecasting result of Mackey-Glass Time Series (6-step-ahead) . . 39
5.4 1-step-ahead Forecasting Result for Canadian Dollar/US Dollar . 41
5.5 1-step-ahead Forecasting Result for Danish Kroner/US Dollar . . 41
5.6 1-step-ahead Forecasting Result for Japanese Yen/US Dollar . . 44
5.7 1-step-ahead Forecasting Result for Mexican New Pesos/US Dollar 44
5.8 1-step-ahead Forecasting Result for Norwegian Kroner/US Dollar 47
5.9 1-step-ahead Forecasting Result for South African Rand/US Dollar 47
5.10 1-step-ahead Forecasting Result for Swiss Francs/US Dollar . . . 50
5.11 1-step-ahead Forecasting Result for US Dollar/Australian Dollar 52
5.12 1-step-ahead Forecasting Result for US Dollar/Euro . . . . . . . 54
5.13 1-step-ahead Forecasting Result for US Dollar/New Zealand Dollar 54
5.14 1-step-ahead Forecasting Result for US Dollar/British Pound . . 57
5.15 Comparison of Historical Volatility . . . . . . . . . . . . . . . . . 58

5.16 Japanese Yen/US Dollar . . . . . . . . . . . . . . . . . . . . . . . 59
5.17 US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . 60
5.18 US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . 63
5.19 Canadian Dollar/US Dollar . . . . . . . . . . . . . . . . . . . . . . 64
5.20 Japanese Yen/US Dollar . . . . . . . . . . . . . . . . . . . . . . . 65
5.21 US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . 65
5.22 Canadian Dollar/US Dollar . . . . . . . . . . . . . . . . . . . . . . 66
5.23 Japanese Yen/US Dollar . . . . . . . . . . . . . . . . . . . . . . . 67
5.24 US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . 68
5.25 Japanese Yen/US Dollar . . . . . . . . . . . . . . . . . . . . . . . 71
5.26 US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . 72
5.27 US Dollar/Australian Dollar . . . . . . . . . . . . . . . . . . . . . 73
5.28 US Dollar/Euro . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75
5.29 Forecasting Results for Canadian Dollar/US Dollar . . . . . . . . . 77
5.30 Forecasting Results for US Dollar/British Pound (1) . . . . . . . . 79
5.31 Forecasting Results for US Dollar/British Pound (2) . . . . . . . . 80

A.1 Middle-term and Long-term Forecasting Results for Canadian Dollar/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . 85
A.2 Middle-term and Long-term Forecasting Results for Danish Kroner/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . 86
A.3 Middle-term and Long-term Forecasting Results for Japanese Yen/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
A.4 Middle-term and Long-term Forecasting Results for Mexican New Pesos/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . 88
A.5 Middle-term and Long-term Forecasting Results for Norwegian Kroner/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . 89
A.6 Middle-term and Long-term Forecasting Results for South African Rand/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . 90
A.7 Middle-term and Long-term Forecasting Results for Swiss Francs/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91
A.8 Middle-term and Long-term Forecasting Results for US Dollar/Australian Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 92
A.9 Middle-term and Long-term Forecasting Results for US Dollar/Euro 93
A.10 Middle-term and Long-term Forecasting Results for US Dollar/New Zealand Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . 94
A.11 Middle-term and Long-term Forecasting Results for US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95

# List of Tables


2.1 Mathematical Definitions of the Activation Functions . . . . . . . 12

5.1 1-step-ahead forecasting performance for Mackey-Glass Time Series 36
5.2 1-step-ahead forecasting performance for Mackey-Glass Time Series 37
5.3 6-step-ahead forecasting performance for Mackey-Glass Time Series 37
5.4 6-step-ahead forecasting performance for Mackey-Glass Time Series 38
5.5 A Comparison of Forecasting Performance for Canadian Dollar/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
5.6 A Comparison of Forecasting Performance for Danish Kroner/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
5.7 A Comparison of Forecasting Performance for Japanese Yen/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
5.8 A Comparison of Forecasting Performance for Mexican New Pesos/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
5.9 A Comparison of Forecasting Performance for Norwegian Kroner/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
5.10 A Comparison of Forecasting Performance for South African Rand/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . 48
5.11 A Comparison of Forecasting Performance for Swiss Francs/US Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49
5.12 A Comparison of Forecasting Performance for US Dollar/Australian Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 51
5.13 A Comparison of Forecasting Performance for US Dollar/Euro . 53
5.14 A Comparison of Forecasting Performance for US Dollar/New Zealand Dollar . . . . . . . . . . . . . . . . . . . . . . . . . . 55
5.15 A Comparison of Forecasting Performance for US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
5.16 Experiment settings one . . . . . . . . . . . . . . . . . . . . . 59
5.17 Forecasting performance for Japanese Yen/US Dollar . . . . . . . 60
5.18 Forecasting performance for US Dollar/British Pound . . . . . . 61
5.19 Experiment settings two . . . . . . . . . . . . . . . . . . . . . 61
5.20 Forecasting performance for US Dollar/British Pound . . . . . . 62
5.21 Experiment settings three . . . . . . . . . . . . . . . . . . . . 62
5.22 Forecasting performance for Canadian Dollar/US Dollar, Japanese Yen/US Dollar and US Dollar/British Pound . . . . . . . . . . . 64
5.23 Experiment settings four . . . . . . . . . . . . . . . . . . . . . 66
5.24 Forecasting performance for Canadian Dollar/US Dollar . . . . . 67

5.25 Forecasting performance for Japanese Yen/US Dollar . . . . . . . 68
5.26 Forecasting performance for US Dollar/British Pound . . . . . . 69
5.27 Experiment settings five . . . . . . . . . . . . . . . . . . . . . . . 70
5.28 Forecasting performance for Japanese Yen/US Dollar and US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . . . 70
5.29 Experiment settings six . . . . . . . . . . . . . . . . . . . . . . . 71
5.30 Forecasting performance for US Dollar/Australian Dollar . . . . 72
5.31 Experiment settings seven . . . . . . . . . . . . . . . . . . . . . . 73
5.32 Forecasting performance for US Dollar/Euro . . . . . . . . . . . 74
5.33 Experiment settings eight . . . . . . . . . . . . . . . . . . . . . . 74
5.34 Forecasting performance for Canadian Dollar/US Dollar . . . . . 76
5.35 Experiment settings nine . . . . . . . . . . . . . . . . . . . . . . . 76
5.36 One-step-ahead Forecasting performance for US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78
5.37 Multiple-step-ahead Forecasting performance for US Dollar/British Pound . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78

# Chapter 1

## Introduction

The foreign exchange market is one of the largest markets in the world. The total trading volume in the global currency market is in excess of 3.2 trillion dollars per day [1]. Foreign exchange rates are among the most important prices in international monetary markets. They affect the economic activity, foreign trade, and the distribution of wealth among countries.

Along with the economic globalization, companies with multinational operations and/or investments are inevitably exposed to the risk that foreign exchange rates will fluctuate in the future. Such fluctuations have the potential to either adversely or favorably affect the asset value of a company. A company's ability to compete internationally will, in some degrees, depend on its capability to appropriately and effectively manage the risks derived from multinational operations. Consequently, forecasting foreign exchange rates has significant meaning for multinational companies. Furthermore, as foreign exchange rates play an important role in international investment, the accuracy in forecasting the foreign exchange rate is also a crucial factor for the success of fund managers. However, foreign exchange rates are affected by many highly correlated economic, political and even psychological factors. The interaction of these factors is in a very complex fashion. Therefore, to predict the movement of foreign exchange rates is a knotty business. As Alan Greenspan, the chairman of America's Federal Reserve, once remarked, forecasting exchange rates "has a success rate no better than that of forecasting the outcome of a coin toss. [6]"

The foreign exchange rate time series usually contain the characteristics of high-noise and non-stationary, which make classical statistical methods incompetent. It is thus necessary to adopt more advanced forecasting techniques. Neural networks are a type of nonlinear model that have proved to be effective for time series prediction. Therefore, in this study, we chose neural network as forecasting tool.

## 1.1 Related Work

In the past two decades, considerable research efforts have been focused on forecasting exchange rates using neural networks. In this section, we will briefly review some of the previous work in this area.

Many studies have demonstrated the forecast ability of the traditional neural network models. For instance, [100] applies multilayer feedforward neural networks to forecast Taiwan Dollar/US Dollar exchange rate and the results are better than ARIMA models'. The authors, therefore, conclude that the neural network approach is a competitive and robust method for the forecasting.

Another example is that [51] uses three neural-network-based forecasting models, i.e. Standard Backpropagation (SBP), Scaled Conjugate Gradient (SCG) and Backpropagation with Baysian Regularization (BPR), to forecast the exchange rates of six foreign currencies against Australian dollar. The results show that all the NN-based models outperform traditional ARIMA model and SCG based model performs best.

However, few studies report opposite results. For example, [83] employs a neural network to study the nonlinear predictability of exchange rates for four currencies at the 1, 6 and 12-step forecasting horizons. The experimental results indicate that the neural network model cannot beat the random walk (RW) in out-of-sample forecast. In other words, the neural networks are lack of the ability of forecasting exchange rate movements.

Fuzzy logic and evolutionary algorithms have been used by a number of studies such as [101], [76] and [98]. [101] uses a so-called intelligent business forecaster which is basically a five-layer fuzzy rule-based neural network to forecast the Singapore Dollar/US Dollar exchange rate and the electricity consumption of Singapore. The experimental results indicate that the performance of the proposed system is superior to Backpropagation Neural Networks and Radial-basis Function Neural Networks in terms of both learning speed and forecasting accuracy.

[76] presents a forecasting model based on neural network and genetic algorithm. The results of the comparative experiments show that the new model provides significantly better performance than traditional artificial neural network and statistical time series modelling approaches.

[98] focuses on how to use evolutionary algorithms to automatically build a radial basis function neural networks for exchange rates forecasting. The results obtained in this study show an improvement of the performance over the conventional methods.

Theoretically, a hybrid model can take advantage of the strong points of both models. Therefore, some studies use hybrid model to forecast exchange rates.

[85] proposes a GRANN-ARIMA hybrid model which combines non-linear Grey Relational Artificial Neural Network and linear ARIMA model for time series forecasting. The experimental results indicate that the hybrid method outperforms ARIMA, Multiple Regression, GRANN, MARMA, MR ANN, ARIMA and traditional Neural Network approaches.

[112] presents a hybrid forecasting model that combines both ARIMA and neural network approaches. The empirical results suggest that the hybrid model outperforms each of the two individual models.

Traditional learning approaches have difficulty with noisy, non-stationary time series prediction. For this reason, [16] uses a hybrid model which combines symbolic processing and recurrent neural networks to solve the problem. The model converts the time series into a sequence of symbols using a self-organizing map(SOM) and then uses recurrent neural networks to perform forecasting. The authors argue that their model can generate useful rules for non-stationary time series forecasting. But as they did not compare the performance of the proposed method with other machine learning approaches, we do not know how effective their method is.

Hybrid models do not always give better performance than individual models. For instance, [91] presents a Neural Net-PMRS hybrid model for forecasting exchange rates. The model uses a traditional multilayer neural network as predictor, but rather than using the last $n$ observations for prediction, the input to the network is determined by the output of the PMRS (Pattern Modelling and Recognition System). The authors compared the performance of the hybrid model with neural networks and PMRS methods and found that there is no absolute winner on all performance measures.

Some studies use economical meaningful variables to improve the forecasting performance of the model. For example, [12] integrates a microeconomic variable (order flow) and some macroeconomic variables including interest rate and crude oil price into a recurrent neural network forecasting model to predict the movements of Canada/US dollar exchange rate. The results demonstrate that the new model provides better forecasting performance than regression models, and both macroeconomic and microeconomic variables are useful for exchange rate forecasting.

A wide variety of other methods have been proposed for forecasting exchange rates: [97] compares the forecasting ability of three different recurrent neural network architectures and then devises a trading strategy based on the forecasting results to optimize the profitability of the system. The author suggests that the recurrent neural networks are particularly suitable for forecasting foreign exchange rates.

In [108], a Flexible Neural Tree(FNT) model is used for forecasting three major international currency exchange rates, namely US Dollar/Euro, US Dollar/British Pound and Japanese Yen/US Dollar. The authors demonstrated that the proposed model provides better forecasts than multilayer feedforward neural network model and adaptive smoothing neural network model.

[55] proposes a multistage neural network meta-learning method for forecasting US Dollar/Euro exchange rate. The empirical results show that the proposed model outperforms the individual ARIMA, BPNN, SVM models and some hybrid models.

[31] examines the use of Neural Network Regression (NNR) models in forecasting and trading the US Dollar/Euro exchange rate. The authors demonstrated that NNR gives batter forecasting and trading results than ARMA models.

In [50], the authors reduce the size of neural networks by using multiple correlation coefficients, principal component analysis and graphical analysis techniques. The experimental results show that pruned neural networks give good forecasting results and the long-term dynamic properties of the resulting neural network models compare favorably with ARIMA models.

Wavelet neural network have been widely-used for forecasting oil prices [90], stock index [9], electricity demand [34] [110] [111] and other time series. But, there are not many applications of wavelet neural network for exchange rate forecasting. [95] is an example, in this study, the author proposes a hybrid model which combines Wavelet Neural Network and Genetic Algorithm for forecasting exchange rates. The experimental results show that the proposed method provides satisfactory performance for different forecasting horizons and, strangely, the author claims that the accuracy of forecasting does not decline when the forecasting horizon increases.

Of course, it is impossible to cover all previous research in this section, but here we provide three useful references: [63] and [99] are two surveys on the applications of neural networks in foreign exchange rates forecasting and a more comprehensive review of the the past 25 years' research in the field of time series forecasting can be found in [37].

From our point of view, the wavelet neural network is especially suitable for forecasting exchange rates, because wavelets can "decompose economic time series into their time scale components" and this is "a very successful strategy in trying to unravel the relationship between economic variables [84]." Furthermore, previous research such as [12] and [38] has shown that using economically significant variables for inputs can improve the performance of the forecasting model. Therefore, we came up with an idea of combining wavelet neural network with economically meaningful variables to achieve accurate forecasting.

The rest of the thesis is organized as follows: Chapter 2 provides a brief introduction to neural networks. Chapter 3 introduces different wavelet techniques.

Chapter 4 explains the design of the proposed forecasting model in detail and the experimental results are presented and discussed in Chapter 5. Finally, Chapter 6 is dedicated to conclusion and future work.

# Chapter 2

## Neural Network

In this chapter, we will provide a brief introduction to the neural network approach. We will start with a general description of the artificial neuron model and then introduce the topology and the training of the neural networks.

## 2.1 Processing Unit

A neural network is a set of interconnected neural processing units that imitate the activity of the brain. These elementary processing units are called neurons. Figure 2.1 illustrates a single neuron in a neural network.

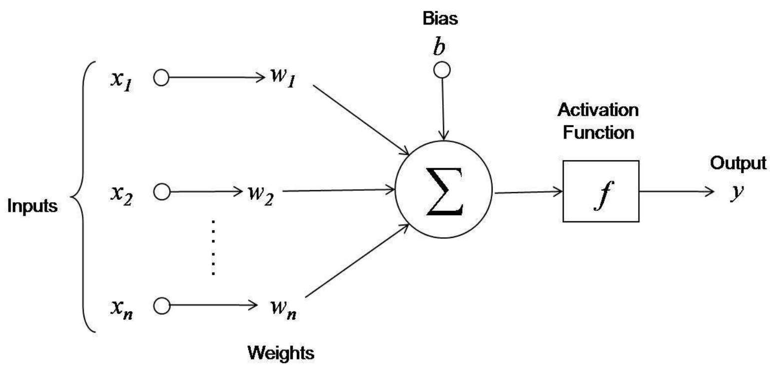



Figure 2.1: Artificial Neuron Model

In the figure, each of the inputs $x_i$ has a weight $w_i$ that represents the strength of that particular connection. The the sum of the weighted inputs and the bias $b$ is input to the activation function $f$ to generate the output $y$. This process can be summarized in the following formula:

$$y = f(\sum_{i=1}^{n} x_i w_i + b) \tag{2.1}$$

A activation function controls the amplitude of the output of the neuron. Neural Networks supports a number of activation functions: a linear activation function (i.e.identity function) returns the same number as was fed to it. This is equivalent to having no activation function. A log-sigmoid activation function (sometimes called unipolar sigmoid function) squashes the output to the range between 0 and 1. This function is the most widely used sigmoid function. A hyperbolic tangent activation function (also called bipolar sigmoid function) is similar to a log-sigmoid function, but it generates outputs between -1 and 1. A symmetric saturating linear function is a piecewise linear version of sigmoid function which provides output between -1 and 1. And a hard Limit function converts the inputs into a value of 0 if the summed input is less than 0, and converts the inputs into 1 if the summed input is bigger than or equal to 0. Figure 2.2 shows these activation functions and their mathematical definitions can be found in Table 2.1.

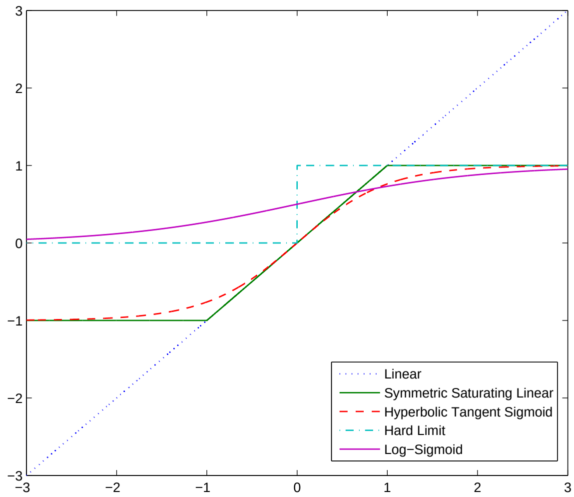



Figure 2.2: Activation Functions

Table 2.1: Mathematical Definitions of the Activation Functions

| Function | Definition | Range |
|---|---|---|
| Linear | $f(x) = x$ | (-inf,+inf) |
| Symmetric Saturating Linear | $f(x) = \begin{cases} -1 & x < -1 \\ x & -1 \leqslant x \leqslant +1 \\ +1 & x > +1 \end{cases}$ | [-1,+1] |
| Log-Sigmoid | $f(x) = \frac{1}{1+e^{-x}}$ | (0,+1) |
| Hyperbolic Tangent Sigmoid | $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ | (-1,+1) |
| Hard Limit | $f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geqslant 0 \end{cases}$ | [0,+1] |

## 2.2 Neural Network Topologies

Generally, neural networks can be classified into two categories based on the pattern of connections: feedforward neural networks and recurrent neural networks. In feedforward neural networks, data move forward from input nodes to output nodes in only one direction. There are no cycles or loops in the network. While in recurrent neural networks, the connections between units form a directed cycle. In this study, we focus on feedforward neural networks.

Early single layer feedforward neural networks contain one input layer and one output layer, the inputs are fed directly to the outputs via a series of weights. Due to their simplicity, they cannot solve non-linearly separable problems such as exclusive-or(XOR) [72]. In order to overcome this problem, multilayer feedforward neural networks were developed [86]. Multilayer neural networks contain one or more hidden layers of neurons between the input and output layers. Each neuron in the layer is connected to every neuron in the next layer, so each neuron receives its inputs directly from the previous layer (except for the input nodes) and sends its output directly to the next layer (except for the output nodes). Traditionally, there is no connection between the neurons of the same layer. Figure 2.3 shows an example of a multilayer neural network with one hidden layer.

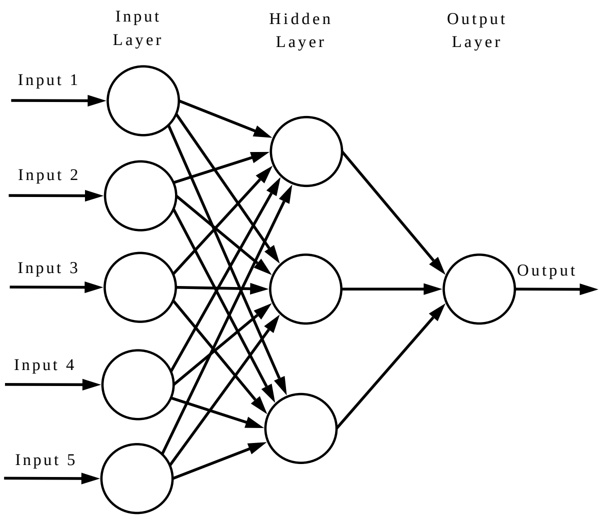



Figure 2.3: Multilayer Feedforward Neural Network

## 2.3 Learning and Generalization

An important feature of neural networks is their ability to learn and generalize. The learning(training) process of a neural network is achieved by adjusting the weights associated with the connections between the neurons. Roughly speaking, there are three types of learning used in neural networks: supervised, unsupervised and reinforcement learning. In supervised learning, a neural network is provided with both inputs and the corresponding desired outputs, the network learns to infer the relationship between them. In unsupervised learning, a neural network is presented only with the inputs and the neural network looks for patterns by itself. Reinforcement Learning can be looked at as an intermediate form of the above two kinds of learning. In reinforcement Learning, the environment supplies inputs to the neural network, receives output, and then provides a feedback. The network adjusts its parameters according to the environmental response. This process is continued until an equilibrium state is reached.

Backpropagation [24] is the most commonly used supervised learning algorithm for multilayer feedforward networks. It works as follows: for each example in the training set, the algorithm calculates the difference between the actual and desired outputs, i.e.the error, using a predefined error function. Then the error is back-propagated through the hidden nodes to modify the weights of the inputs. This process is repeated for a number of iterations until the neural network converges to a minimum error solution. Although backpropagation algorithm is a suitable method for neural network training, it has some shortcomings such as slow convergence and easily trapped in local minima. For this reason, several improved learning algorithms are proposed including Scaled Conjugate Gradient(SCG) [75], Levenberg-Marquardt(LM) [92], Resilient Propagation(RProp) [71] and so on.

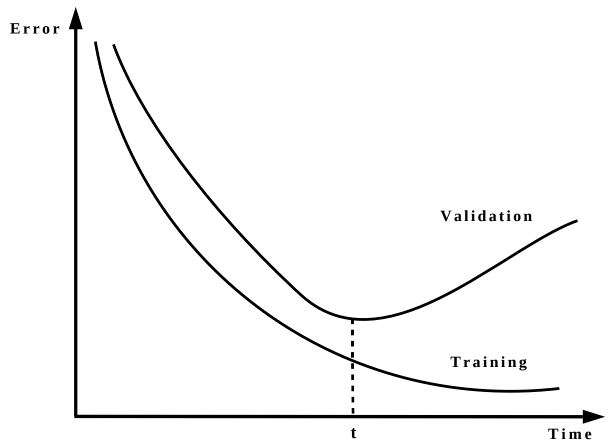



Figure 2.4: Overfitting in Supervised Learning

The performance of the neural network critically depends on its generalization ability. However, a neural network can suffer from either underfitting or overfitting problem. To obtain a better generalization, we have to make a trade-off between underfitting and overfitting. Underfitting means that the neural network performs poorly on new samples because it has not captured the underlying patterns in the training data. This problem can be prevented by adding sufficient hidden nodes to the neural network and training the network for long enough. Overfitting means that the network performs poorly on unseen data as it is too highly-trained to the noise of the training samples. Figure 2.4 illustrates the overfitting problem. We can see that the validation error increases while the training error decreases steadily. There are several ways to avoid overfitting, for instance, we can restrict the number of hidden units in the neural network, and we can stop the training early when overfitting starts to happen as well as use cross-validation technique. However, there is no overall best solution for the overfitting problem. In Section 4.2, we will explain our strategy to ensure the generalization of the neural networks.

# Chapter 3

## Wavelet Transform

*"All this time the guard was looking at her, first through a telescope, then through a microscope, and then through an opera glass."*
*- Lewis Carroll, Through the Looking Glass (chapter3)*

In this chapter, we will briefly introduce different wavelet techniques.

### 3.1 Why Wavelet Transform ?

According to Fourier theory, a signal can be expressed as the sum of a series of sines and cosines. This sum is also called as a Fourier expansion (see Equation 3.1). However, a serious drawback of the Fourier transform is that it only has frequency resolution and no time resolution. Therefore, we can identify all the frequencies present in a signal, but we do not know when they are present. To overcome this problem, the wavelet theory is proposed.

$$\begin{aligned} F(\omega) &= \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \\ &= \int_{-\infty}^{+\infty} f(t)(\cos \omega t - j \sin \omega t) dt \end{aligned} \tag{3.1}$$

Mathematically, a wavelet can be defined as a function with a zero average:

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \tag{3.2}$$

and the basic idea of the wavelet transform is to represent any function as a superposition of a set of wavelets or basis functions. Because these functions are small waves located in different times, the wavelet transform can provide information about both the time and frequency domains.

### 3.2 Continuous Wavelet Transform (CWT)

Similar to equation 3.1, the continuous wavelet transform is defined as:

$$\gamma(s, \tau) = \int_{-\infty}^{+\infty} f(t) \Psi_{s,\tau}^{*}(t) dt \tag{3.3}$$

where $*$ denotes complex conjugation. The variables $s$ and $\tau$ are the new dimensions, i.e. scale and translation, after the wavelet transform. Equation 3.3 shows how a function $f(t)$ is decomposed into a set of basis functions $\Psi_{s,\tau}(t)$ which is call the wavelets. And the wavelets are generated from a so-called mother wavelet by scaling and translation:

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \tag{3.4}$$

where $s$ is a scale factor that reflects the scale i.e. width of a particular basis function, $\tau$ is a translation factor that specifies the translated position along the $t$ axis and the factor $\frac{1}{\sqrt{s}}$ is for energy normalization across the different scales. Just as Fourier transform decomposes a signal into a series of sines and cosines with different frequencies, continuous wavelet transform decomposes a signal into a series of wavelets with different scales $s$ and translations $\tau$.

## 3.3 Discrete Wavelet Transform (DWT)

The disadvantage of the continuous wavelet transform lies in its computational complexity and redundancy. In order to solve these problems, the discrete wavelet transform is introduced. Unlike CWT, the DWT decomposes the signal into mutually orthogonal set of wavelets. The discrete wavelet is defined as:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}}\psi\left(\frac{t-k\tau_0 s_0^j}{s_0^j}\right) \tag{3.5}$$

where $j$ and $k$ are integers, $s_0 > 1$ is a fixed dilation step and the translation factor $\tau_0$ depends on the dilation step. The scaling function and the wavelet function of DWT are defined as:

$$\varphi(2^j t) = \sum_{i=1}^{k} h_{j+1}(k)\varphi(2^{j+1}t-k) \tag{3.6}$$

$$\psi(2^j t) = \sum_{i=1}^{k} g_{j+1}(k)\varphi(2^{j+1}t-k) \tag{3.7}$$

And then, a signal $f(t)$ can be written as:

$$f(t) = \sum_{i=1}^{k} \lambda_{j-1}(k)\varphi(2^{j-1}t-k) + \sum_{i=1}^{k} \gamma_{j-1}(k)\psi(2^{j-1}t-k) \tag{3.8}$$

The discrete wavelet transform can done by using the filter bank scheme developed by [68]. Figure 3.1 shows a two-channel filter bank scheme for DWT.

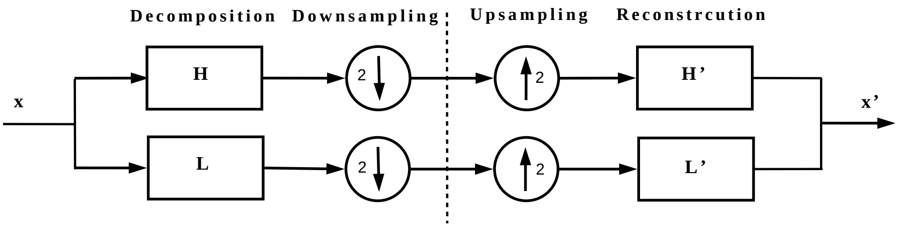



Figure 3.1: Filter Bank Scheme for DWT

In the figure, H,L,and H',L' are the high-pass and low-pass filters for wavelet decomposition and reconstruction respectively. In the decomposition phase, the low-pass filter removes the higher frequency components of the signal and high-pass filter picks up the remaining parts. Then, the filtered signals are down-sampled by two and the results are called approximation coefficients and detail coefficients. The reconstruction is just a reversed process of the decomposition and for perfect reconstruction filter banks, we have $x = x'$. A signal can be further decomposed by cascade algorithm as shown in Equation 3.9:

$$\begin{aligned} x(t) &= A_1(t) + D_1(t) \\ &= A_2(t) + D_2(t) + D_1(t) \\ &= A_3(t) + D_3(t) + D_2(t) + D_1(t) \\ &= A_n(t) + D_n(t) + D_{n-1}(t) + ...... + D_1(t) \end{aligned} \tag{3.9}$$

where $D_n(t)$ and $A_n(t)$ are the detail and the approximation coefficients at level $n$ respectively. Figure 3.2 illustrates the corresponding wavelet decomposition tree.

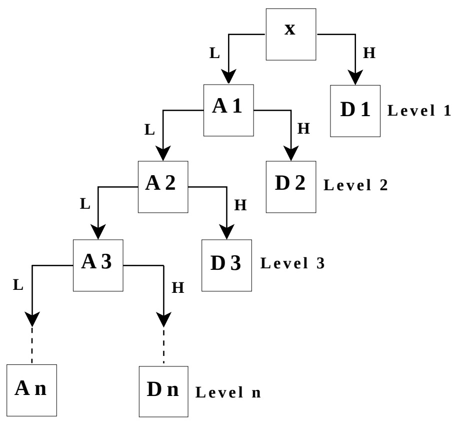



Figure 3.2: Wavelet Decomposition Tree

A significant potential problem with the DWT is that it is a shift variant transform. Shift-variance is a "phenomenon of not necessarily matching the shift of the one-level DWT with the one-level DWT of the same shift of a data sequence [22]". Due to shift-variance, small shifts in the input waveform may cause large changes in the wavelet coefficients and this makes the DWT unsuitable for this study because we cannot "relate the information at a given time point at the different scales [78]". Therefore, we need a shift-invariant or at least time-invariant wavelet transform for our forecasting model. (For a detailed discussion of the shift-variance problem, see [22] [26] [15] [78])

## 3.4 Stationary Wavelet Transform (SWT)

Previous studies [15] have pointed that the shift-variance results from the downsampling operation in the DWT(see Figure 3.1). Hence, the simplest way to avoid shift-variance is to skip the downsampling operation, and this is, in fact, the main difference between the SWT and DWT. Figure 3.3 shows the filter bank of the SWT. We can see that the SWT is similar to the DWT except that in SWT, the signal is never sub-sampled and instead, the signal are upsampled at each level of decomposition. Therefore, SWT is shift-invariant and this is exactly what we need. Figures 3.5 and 3.6 shows the results of the stationary wavelet decomposition and discrete wavelet decomposition for Japanese Yen/US Dollar time series. We can observe a noticeable difference.

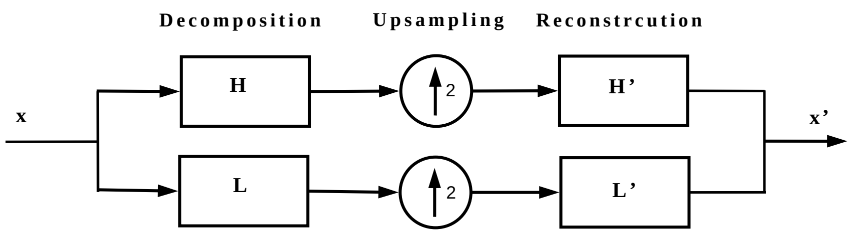



Figure 3.3: Filter Bank Scheme for SWT

## 3.5 Wavelet Packet Transform

The structure of wavelet packet transform is very similar to discrete wavelet transform. The difference is that in the discrete wavelet transform, only the approximation coefficients are decomposed, while in the wavelet packet transform, both the detail and approximation coefficients are decomposed. Therefore, wavelet packet transform offers a more complex and flexible analysis. Figure 3.4 shows a wavelet packet decomposition tree over 3 levels. we can see that a $n$-level wavelet packet decomposition produces $2^n$ different sets of coefficients as opposed to $n+1$ sets in the discrete wavelet transform.

## 3.6 Wavelet and Wavelet Packet Denoising

Denoising is to remove noise as much as possible while preserving useful information as much as possible. The basic noisy signal model [40] has the following

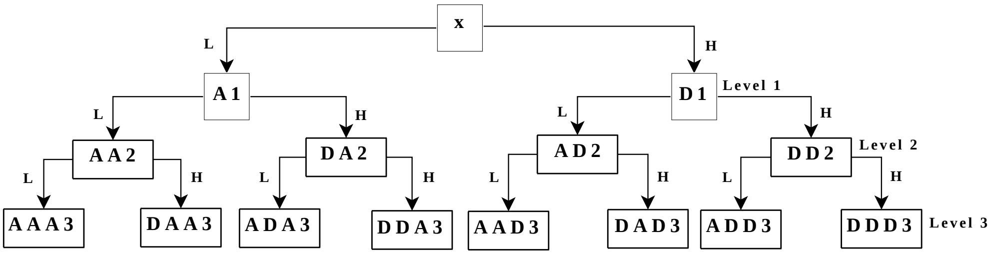



Figure 3.4: Wavelet Packet Decomposition Tree

form:

$$s(x) = f(x) + e(x) \tag{3.10}$$

where $s(x)$ is the observed signal, $f(x)$ is the original signal, $e(x)$ is Gaussian white noise with zero mean and variance $\sigma^2$. The objective of denoising is to suppress the noise part of the signal $s$ and to recover $f$.

The basic ideas and procedures of wavelet and wavelet packet denoising (also referred to as wavelet shrinkage) are exactly the same: as the noise in a signal is mostly contained in the details of wavelet coefficients, that is, the high frequency range of the signal [52], if we set the small coefficients to zero, much of the noise will disappear and of course, inevitably, some minor features of the signal will be removed as well. The denoising procedure can be done in three steps: 1. Select a wavelet and a level $n$, apply wavelet/wavelet packet decomposition to the noisy signal to produce wavelet coefficients. 2. For each level from 1 to $n$, choose a threshold value and apply thresholding to the detail coefficients. 3. Perform wavelet/wavelet packet reconstruction to obtain a denoised signal.

The most widely-used thresholding methods are hard-thresholding:

$$T_\lambda(x) = \begin{cases} 0 & if\,|x| \leqslant \lambda \\ x & otherwise \end{cases} \tag{3.11}$$

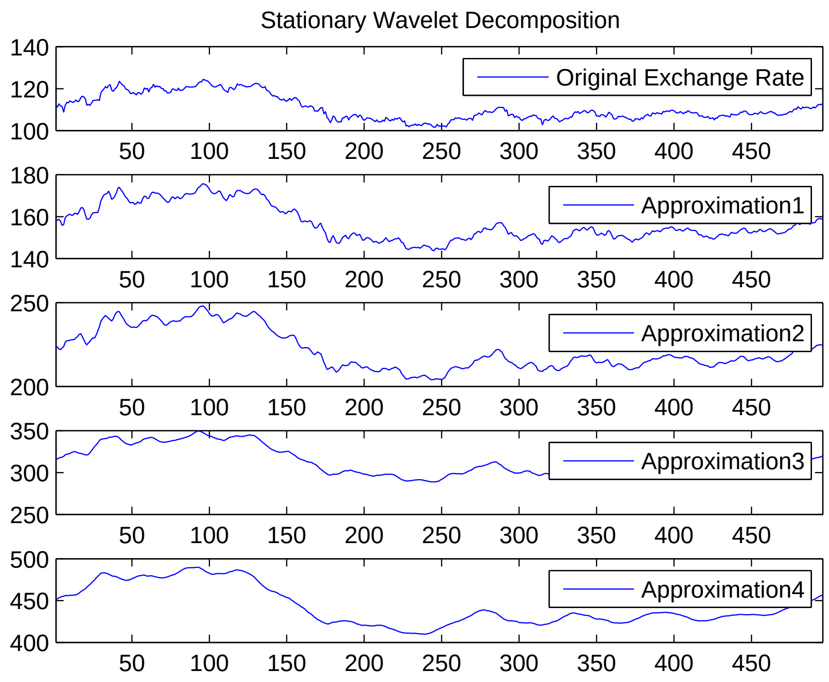
and soft-thresholding [28] [29]:

$$T_\lambda(x) = \begin{cases} x - \lambda & if\ x > \lambda \\ 0 & if\,|x| \leqslant \lambda \\ x + \lambda & if\ x < -\lambda \end{cases} \tag{3.12}$$

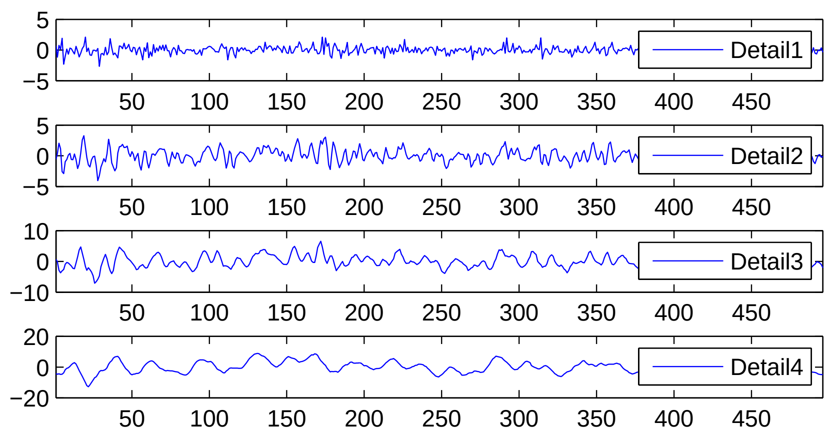
where $\lambda$ can be calculated by Stein's Unbiased Risk Estimate(SURE) method:

$$\lambda = \sqrt{2\log_e(n\log_2(n))} \tag{3.13}$$

where $n$ is the length of the signal. In this study, we used the soft-thresholding approach, because it has been reported that the soft-thresholding is more effective than the hard-thresholding [36] [93].

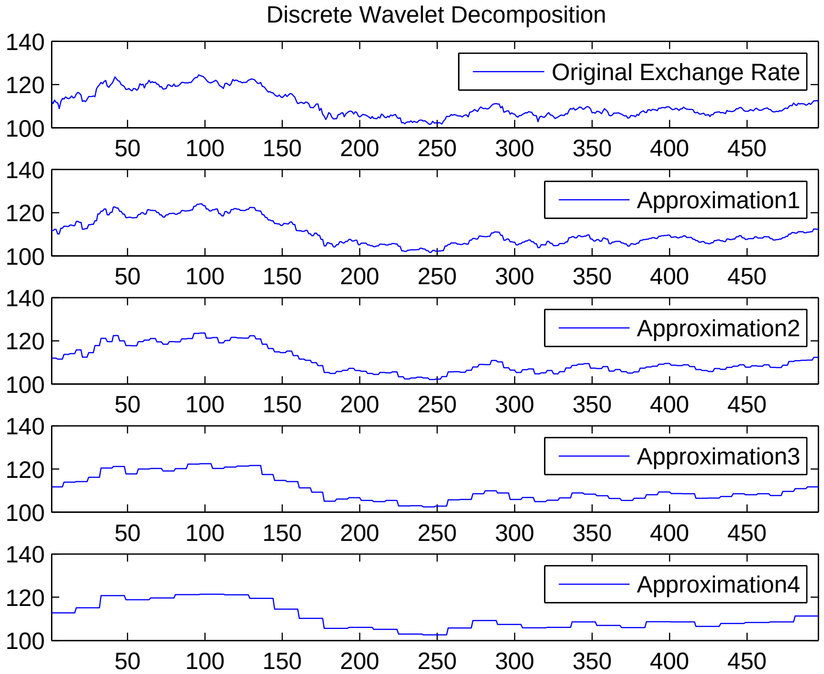



(a) Original Time Series and Approximations

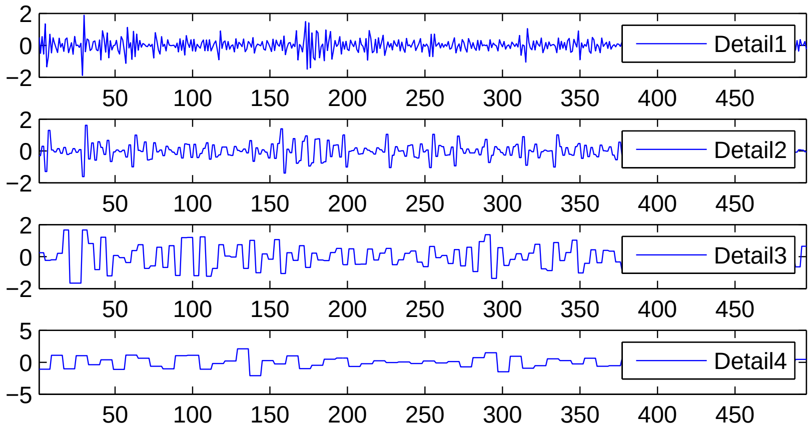



(b) Details

Figure 3.5: Level 4 Stationary Wavelet Decomposition using Haar Wavelet




(a) Original Time Series and Approximations




(b) Details

Figure 3.6: Level 4 Discrete Wavelet Analysis using Haar Wavelet

# Chapter 4

## Forecast Model

In this chapter, we will introduce the design of the proposed forecasting model in detail.

### 4.1 Overview of the Design

In this study, we implemented four different forecasting models including pure neural network forecasting model, wavelet and wavelet-packet denoising based models as well as stationary wavelet transform based forecasting model.

A pure neural network forecasting model (see Figure 4.1) takes raw time series as input signal to predict the future. Such a forecasting model is easy to implement. However, the noise in the time series data will seriously affect the accuracy of the forecast. In order to solve the problem, wavelet denoising based model is proposed. Figure 4.2 illustrates a typical wavelet denoising based forecasting model in which the time series raw data is pre-processed using wavelet denoising technique before being forwarded to the neural network predictor.

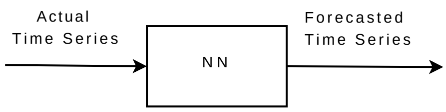



Figure 4.1: Pure Neural Network Forecasting Model

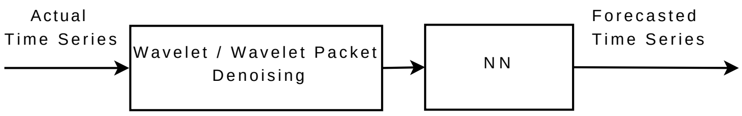



Figure 4.2: Wavelet/Wavelet Packet Denoising-based Forecasting Model

Figures 4.3 and 4.4 show our hybrid wavelet-neural network scheme for time series prediction. It consists of two independent models: one for training and another for forecasting, and in the system, the time series data are divided into three parts: training set, validation set and testing set.

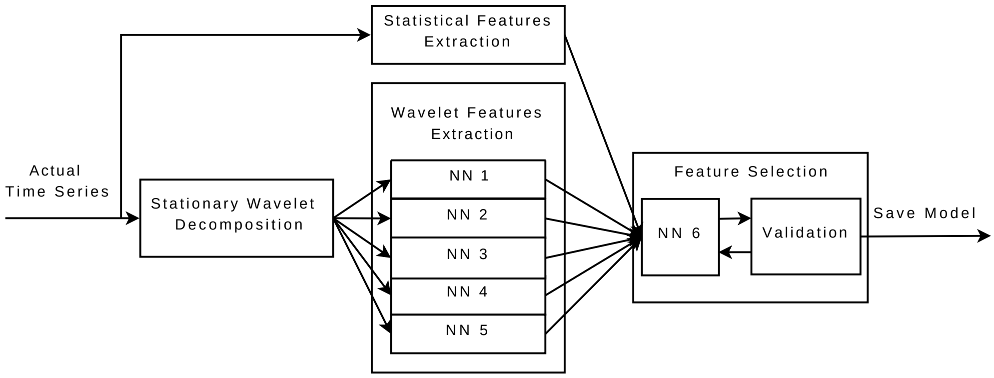



Figure 4.3: Proposed Model: Training

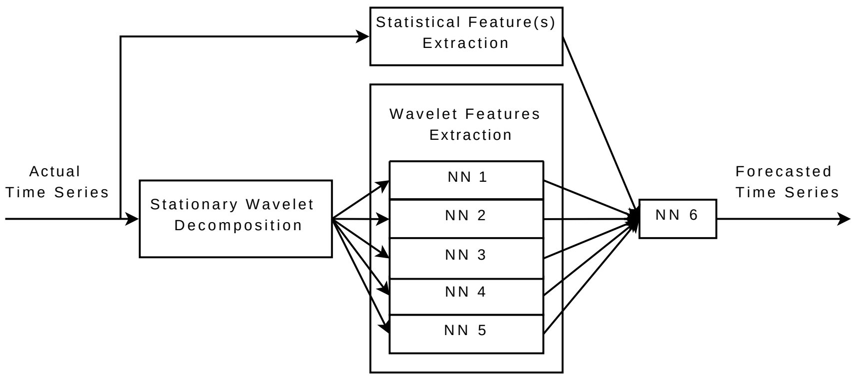
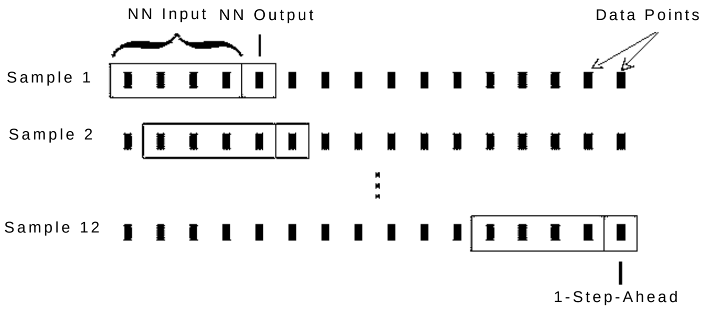
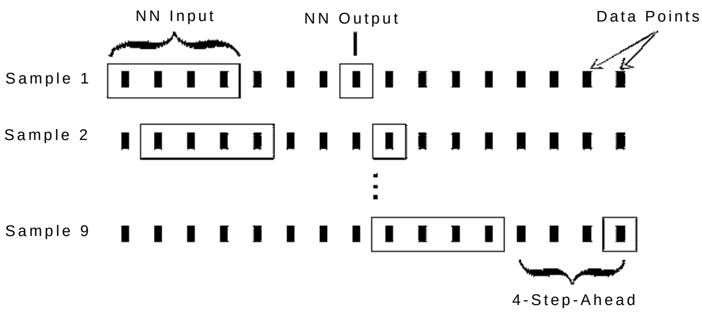



Figure 4.4: Proposed Model: Forecasting

The training model basically involves three stages, data preprocessing, feature extraction and feature selection. More concretely, given a time series $f(x)$, $x = 1, 2, 3...n$, the task is to forecast $p$ step ahead i.e.$f(x + p)$ and the size of the time-window is $q + 1$. The output of the forecasting model is suppose to be a function of the values of the time series within the time-window $[f(x), f(x - 1), ..., f(x - q)]$, $x - q \geqslant 1$ (see Figure 4.5). In the preprocessing phase, the training set are decomposed at level 4 using stationary wavelet analysis to obtain the 4 high frequency details and 1 low frequency approximation (see Equation 3.9 and Figure 3.2). In the second phase, the results of the wavelet decomposition are fed to 5 neural network predictors, i.e. NN1 to NN5, to generate 5 wavelet features. In this step, NN1 to NN5 are trained to forecast the values of $f(x + p)$ for each decomposition scale separately and at the same time, 8 statistical features (see Section 4.4) are extracted from the original time series. In the third stage, we train NN6 using exhaustive feature selection strategy to find the combination which provides the best forecast for the validation set, i.e. the smallest RMSE, and save the trained model. The forecasting model has a similar structure as the training model except the validation component. It utilizes the trained neural networks to forecast $f(x+p)$ directly. Therefore, the number of the inputs of NN1 to NN5 is $q+1$, while NN6 has $5+m$ inputs, here $m$ is the number of the statistical features that are selected in the training phase. The number of the output of each neural network in the model is 1. The choice of hidden layers and hidden units will be explained in Section 4.2




(a) One-Step-Ahead Forecasting*




(b) Multiple-Step-Ahead Forecasting*

Figure 4.5: An Illustration of Different Forecasting Modes
* This figure is modified from [81]

We implemented the four forecasting models in Matlab R2008b with wavelet toolbox [3]. The neural networks in the system were created and trained using Netlab toolbox [5] and the statistical features were extracted using NaN-tb statistic toolbox [4]. The neural networks in the system are trained using Scaled Conjugate Gradient (SCG) algorithm and the activation function of output layer is linear. All the input data are normalized to the range [-1,1] before being fed into the neural networks.

## 4.2 Network Structure Design

To define a network architecture, we have to decide how many hidden layers should be used and how many hidden units should be in a hidden layer. In this section, we will explain our design.

The choice of the number of hidden layers is an important issue because too many hidden layers can degrade the network's performance. Theoretically, neural networks with two hidden layers can approximate any nonlinear function to an arbitrary degree of accuracy. Hence, there is no reason to use neural networks with more than two hidden layers. Furthermore, it has also been proved that for the vast majority of practical problems, there is no need to use more than one hidden layer. Therefore, all the neural networks in our model only have one hidden layer.

Another crucial issue is the choice of optimal number of hidden units in neural networks. If there are too few hidden nodes, the system will get high training error and high generalization error due to underfitting. On the contrary, if the number of hidden units is too large, then the model may get low training error but still have high generalization error due to overfitting, as "attempting a large number of models leads to a high probability of finding a model that fits the training data well purely by chance" [27] [46]. In our design, we follow Occam's razor principle, i.e. select the simplest model which describes the data adequately. More concretely, for each neural network in the system, we set the number of hidden neurons by using half the sum of inputs plus output, which is a rule-of-thumb that has been used in previous research such as [110] [111].

## 4.3 Wavelet Features

In the proposed model, we use Stationary Wavelet Transform(SWT) to extract wavelet features. But, we actually tested both the SWT and DWT versions of the model and the results will be presented in Chapter 5. The reason why we did this is because although shift-variance is a well-known disadvantage of the discrete wavelet transform, some studies such as [90] [102] [57] still suffer from this problem. Therefore, we think that it is necessary to show the impact of using non-shift-invariant discrete wavelet transform on the forecasting performance.

In this study, we determined the wavelet decomposition levels and wavelet functions empirically. We found that a wavelet decomposition at level 4 provides the best forecasting results for the proposed model while a resolution at level 2 is more suitable for the wavelet-denoising-based methods.

We chose Haar wavelet as wavelet function for our model(SWT version) because it gives best forecasting performance. Figure 4.6 shows the wavelet function and the scaling wavelet function of the Haar wavelet and the corresponding mathematical definitions are given in Equations 4.1 and 4.2.

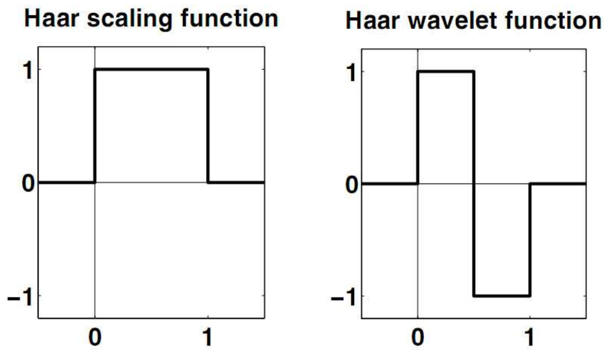



Figure 4.6: The Haar Wavelet*
* This figure is taken from [52]

$$\psi(x) = \begin{cases} 1 & 0 \leqslant x < \dfrac{1}{2} \\ -1 & \dfrac{1}{2} \leqslant x < 1 \\ 0 & otherwise \end{cases} \tag{4.1}$$

$$\phi(x) = \begin{cases} 1 & 0 \leqslant x < 1 \\ 0 & otherwise \end{cases} \tag{4.2}$$

When we selected wavelet function for the wavelet-denoising-based model and the DWT version of the proposed model, we found that the forecasting performance of the models is proportional to the order of the wavelet and the Db40 wavelet works best. Therefore, we use this wavelet in these models. Figure 4.7 shows the wavelet function and the scaling wavelet function of the Db40 wavelet.

Furthermore, we observed that for the low-volatility part a time series, the difference of the performance between wavelet and wavelet packet denoising is not remarkable. However, for the high-volatility part of the data, wavelet-packet denoising can, sometimes, provide a much better denoising effect. Figure 4.8 shows a comparison of the denoising performance of the wavelet and wavelet-packet denoising for US Dollar/Euro time series. We can see that wavelet-packet denoising technique can remove noise from raw time series without smoothing out the peaks and bottoms. It means that, in this case, wavelet-packet denoising preserves more information of the original signal.

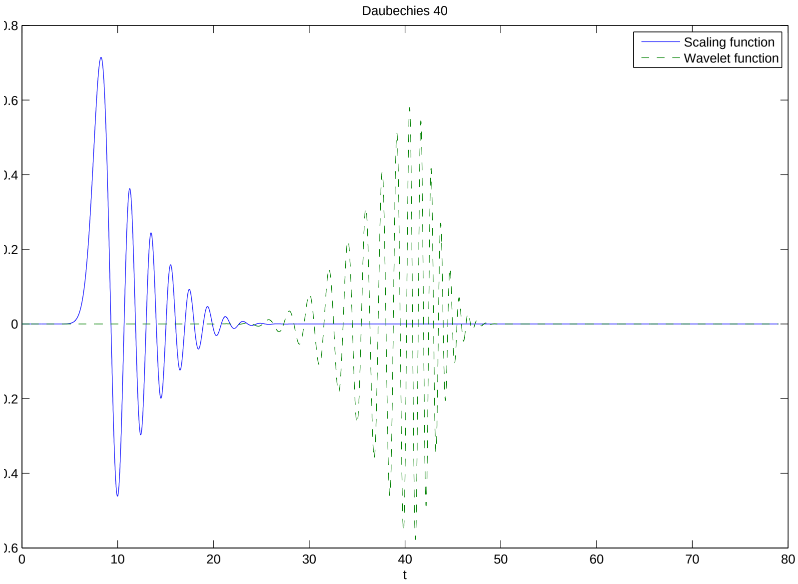



Figure 4.7: The Db40 Wavelet

## 4.4 Statistical Features

Besides wavelet features, we employed eight additional features to represent the characteristics of time series. Five of them, i.e. Mean, Mean Absolute Deviation, Variance, Skewness and Kurtosis, are basic descriptive statistical measures that are commonly-used in economics and finance. Examples of the application of these statistical measures can be found in [62] [59] [89] [79] [19] [58] [39]. In the proposed model, we use them to describe the data inside the time-window. The rest three features i.e. Turning Points, Shannon Entropy and Historical Volatility have also been proved to be particularly useful for financial analysis.

### 4.4.1 Mean

The Mean, traditionally denoted by $\bar{x}$, is the arithmetic average of a set of observations. It can be calculated by the sum of the observations divided by the number of observations:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.3}$$

where $x_i$ denotes the value of observation $i$ and $n$ is the number of observations.

### 4.4.2 Mean Absolute Deviation

The mean absolute deviation is a measure of the variability. For a sample size $n$, the mean absolute deviation is defined by:

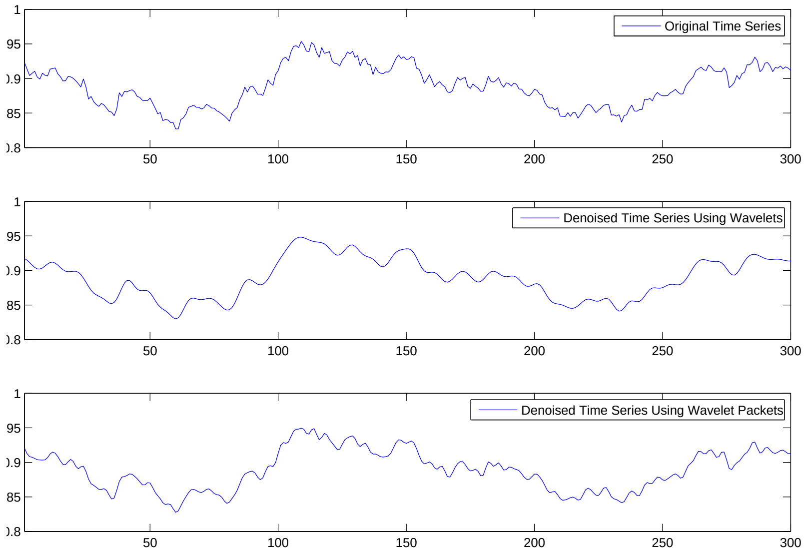



Figure 4.8: A Comparison of the Wavelet and Wavelet-Packet Denoising Techniques (Db40 wavelet function is used)

$$Mad(x_1....x_n) = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| \tag{4.4}$$

where $\bar{x}$ is the mean of the distribution.

### 4.4.3 Variance

The variance is another widely-used measure of dispersion. It estimates the mean squared deviation of $x$ from its mean value $\bar{x}$:

$$Var(x_1...x_n) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{4.5}$$

### 4.4.4 Skewness

The skewness measures the degree of asymmetry of a distribution around the mean. In other words, the skewness tells us about the direction of variation of the data set: positive skewness indicates that the tail of the distribution is more stretched on the side above the mean; negative skewness signifies that the tail of the distribution extends out towards the side below the mean; the skewness of a normal distribution is zero. The conventional definition of the skewness is:

$$Skew(x_1...x_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{\sigma}\right]^3 \tag{4.6}$$

where $\sigma = \sigma(x_1...x_n) = \sqrt{Var(x_1...x_n)}$ is the standard deviation of the distribution.

### 4.4.5 Kurtosis

The kurtosis measures the relative peakedness or flatness of a distribution: positive kurtosis indicates a peaked distribution (i.e. leptokurtic); negative kurtosis means a flat distribution (i.e. platykurtic); the kurtosis of a normal distribution (i.e. mesokurtic) is zero. The kurtosis can be defined as follows:

$$Kurt(x_i...x_n) = \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{\sigma}\right]^4\right\} - 3 \tag{4.7}$$

### 4.4.6 Turning Points

Mathematically speaking, a turning point can be defined as a local maxima or minima of a time series. More precisely, a time series $x$ has a turning point at time $t$, if

$$(x_{t+1} - x_t)(x_t - x_{t-1}) < 0 \tag{4.8}$$

where $1 < t < n$. Turning Points are important indictors of the economic cyclical behavior that has been used in some studies such as [44], [49], [104] and [33]. [41] suggests that the counts and proportions of turning points can be directly related to the data generating process and, therefore, provides a methodology for modelling economic dynamics. In this study, we employ the frequency (i.e. the numbers) of turning points observed within the time-window as a feature to describe time series data. A general review about the application of turning points in economics can be found in [61] and [10].

### 4.4.7 Shannon Entropy

The entropy of a random variable is defined in terms of its probability distribution and can be used as a measure of randomness or uncertainty. Let $X$ be a discrete random variable taking a finite number of possible values $x_1, x_2, ..., x_n$ with probabilities $p_1, p_2, ..., p_n$ respectively such that $p_i \geqslant 0, i = 1, 2, ..., n, \sum_{i=1}^{n} p_i = 1$. Shannon entropy is then defined by:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{4.9}$$

Shannon entropy has been utilized for time series analysis by a number of studies [74] [73] [23] [88] [70] [87]. When applied for time series forecasting task [109], Shannon entropy measures the predictability of future values of the time series based on the probability distribution of the values already observed in the data. In our implementation, we calculate the Shannon entropy i.e. the average uncertainty of the variables in the time-window.

### 4.4.8 Historical Volatility

The historical volatility is a measure of price fluctuation over a given time period. It uses historical price data to empirically estimate the volatility of a financial instrument in the past. Instruments that have large and frequent price movements are said to be of high volatility; instruments whose price movements are predictable or slow are considered to be low volatile. The historical volatility is widely-used as an indicator of risk in finance [77] [35] [13] [113] [18] [67] [32] [60]. A comprehensive review can be found in [82].

The historical volatility can be calculated by Exponentially Weighted Moving Average (EWMA) method:

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)y_{n-1}^2 \tag{4.10}$$

where $\sigma_n$ is the estimation of the volatility for day $n$ based on the calculation at the end of day $n-1$, $\lambda$ is a decay factor between 0 and 1, $\sigma_{n-1}$ is the volatility estimate a day ago and $y_{n-1}$ is the most recently daily observation. According to [48], $\lambda = 0.94$ provides the closest estimate to the real volatility. In the proposed model, we employ a neural network predictor (see Figure 4.1) to predict the movement of the historical volatility time series and then use the the result of the forecasting as a feature.

# Chapter 5

## Experimental Results and Discussions

We conducted a series of comparative analysis to evaluate the performance of our system. Firstly, we tested the system with Mackey-Glass time series which is a well-known benchmark that has been used and reported by a number of researchers. Then the proposed model is used to predict real-world exchange rates time series. And thirdly, we compared the performance of our method with others' work. In this chapter, we will present and discuss the results we obtained.

### 5.1 Mackey-Glass Time Series

The Mackey-Glass time series is generated from the following time-delay differential equation:

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x^{10}(t-\tau)} - bx(t) \tag{5.1}$$

We generate 1000 points with the initial condition of $x(0) = 1.2$, $\tau = 17$, $a = 0.2$, $b = 0.1$ and we take data sets from $t = 123$ to $t = 1123$. The first 500 data points are used for the training and validating, the remaining 500 points are reserved for the testing phase. In 1-step-ahead test, we use $x(t)$, $x(t+1)$, $x(t+2)$ ...... $x(t+9)$ to forecast the value of $x(t+10)$, while in 6-step-ahead test, the task is to predict the $x(t+6)$ using the input variables $x(t)$, $x(t-6)$, $x(t-12)$, $x(t-18)$. Figures 5.1, 5.2 and 5.3 show the original and predicted Mackey-Glass time series respectively, and Tables 5.1, 5.2, 5.3, 5.4 are comparisons of the forecasting performance of various methods. In Table 5.3, non-dimensional error index(NDEI) is defined as RMSE divided by the standard deviation of the target series [45]. It is worth to mention that in [45], the authors presented their experimental results in NDEI. However, surprisingly, many later studies [17] [53] [43] [25] [42] used those results as RMSE. The reason why we do two tests and present the results using different performance measures is because pervious research used different experimental settings. In order to make a fair comparison with them, we have to do the same.

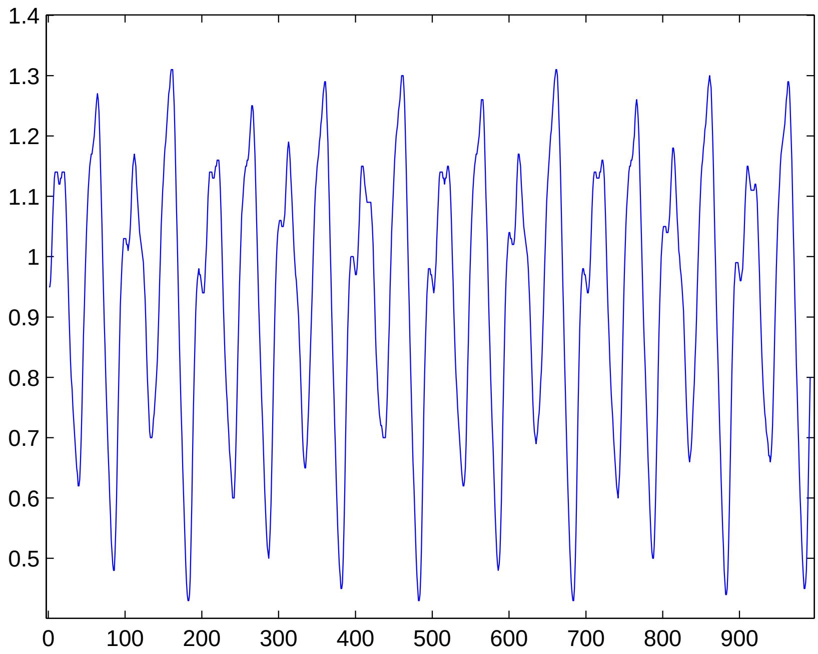



Figure 5.1: Mackey-Glass Time Series

Table 5.1: 1-step-ahead forecasting performance for Mackey-Glass Time Series

| Method | RMSE | Ranking |
|---|---|---|
| Neural tree [106] | 0.0069 | 6 |
| PNN [94] | 0.0059 | 5 |
| NN+Wavelet Denoising | 0.0028 | 4 |
| NN+Wavelet Packet Denoising | 0.0024 | 3 |
| New Method(SWT) | 0.0015 | 2 |
| New Method(DWT) | 0.0002 | 1 |

We can see from Tables 5.1 that the proposed model outperforms other methods in 1-step-ahead test. In Table 5.2, the result from the new model(SWT) is worse than the results from [57]. But, we want to emphasize that [57] used discrete wavelet packet transform with DB2 wavelet and, in fact, the result from the DWT version of our model is much better than theirs. Tables 5.3 and 5.4 show that, in 6-step-ahead test, the proposed model gives reasonable results and we also noticed that wavelet-denoising based approaches provide better forecasting performance than our method. It seems that those approaches have some advantages in certain cases. Whether this is true or not, we need to see the forecasting results obtained from real-world data set.

Table 5.2: 1-step-ahead forecasting performance for Mackey-Glass Time Series

| Method | MSE | Ranking |
|---|---|---|
| NN+Wavelet Denoising | 7.91e-6 | 6 |
| NN+Wavelet Packet Denoising | 5.98e-6 | 5 |
| New Method(SWT) | 2.16e-6 | 4 |
| WP-MLP with hierarchical clustering [57] | 1.53e-6 | 3 |
| WP-MLP with counterpropagation clustering [57] | 2.42e-7 | 2 |
| New Method(DWT) | 4.97e-8 | 1 |

Table 5.3: 6-step-ahead forecasting performance for Mackey-Glass Time Series

| Method | NDEI | Ranking |
|---|---|---|
| Linear Predictive Model [45] | 0.55 | 12 |
| Auto Regressive Model [45] | 0.19 | 11 |
| New Fuzzy Rules [64] | 0.0816 | 10 |
| Cascade Correlation NN [45] | 0.06 | 9 |
| 6th-order Polynomial [45] | 0.04 | 8 |
| New Method(SWT) | 0.0234 | 7 |
| Back Prop NN [45] | 0.02 | 6 |
| New Method(DWT) | 0.0173 | 5 |
| NN+Wavelet Denoising | 0.0169 | 4 |
| NN+Wavelet Packet Denoising | 0.0156 | 3 |
| ClusNet [96] | 0.014 | 2 |
| ANFIS [45] | 0.007 | 1 |

## 5.2 Exchange Rates Forecasting

In foreign exchange trading, more than half of all trading directly involves the exchange of dollars. The importance of trading other currencies against U.S. dollars arises from the fact that currency dealers quote other currencies against the dollar when trading among themselves [2]. Hence, in this study, all foreign exchange rates will be quoted as the number of units of the foreign currency per US dollar. The data sets that we used are obtained from the official website of American Federal Reserve Bank [7]. 11 currency exchange rates are selected including Canadian Dollar/US Dollar, Danish Kroner/US Dollar, Japanese Yen/US Dollar, Mexican New Pesos/US Dollar, Norwegian Kroner/US Dollar, South African Rand/US Dollar, Swiss Francs/US Dollar, US Dollar/Australian Dollar, US Dollar/Euro, US Dollar/New Zealand Dollar and US Dollar/British Pound. We take daily data from to January 1 1999 to January 31 2007 with a total of 2032 observations. To serve different purposes,

Table 5.4: 6-step-ahead forecasting performance for Mackey-Glass Time Series

| Method | RMSE | Ranking |
|---|---|---|
| Wang (product operator) [65] | 0.0907 | 17 |
| Min Operator [53] | 0.0904 | 16 |
| Kim and Kim [25] | 0.026 | 15 |
| Classical RBF [21] | 0.0114 | 14 |
| WNN+gradient [107] | 0.0071 | 13 |
| WNN+hybrid [107] | 0.0059 | 12 |
| New Method(SWT) | 0.0053 | 11 |
| HWNN [105] | 0.0043 | 10 |
| LLWNN+gradient [107] | 0.0041 | 9 |
| New Method(DWT) | 0.0039 | 8 |
| NN+Wavelet Denoising | 0.0038 | 7 |
| LLWNN+hybrid [107] | 0.0036 | 6 |
| new RBF structure [43] | 0.0036 | 6 |
| NN+Wavelet Packet Denoising | 0.0035 | 5 |
| PG-RBF network [42] | 0.00287 | 4 |
| FNT [103] | 0.0027 | 3 |
| Evolving RBF with input selection [30] | 0.00081 | 2 |
| improved ANFIS [17] | 0.00055 | 1 |

the data sets are divided into three parts: the training set covers January 1 1999 to January 31 2005 with 1529 observations, the validation set runs from February 1 2005 to January 31 2006 with 251 observations while the testing set is from February 1 2006 to January 31 2007 with 252 observations.

Generally speaking, exchange rate forecasting problems can be classified into three categories according to different forecast horizons : "short-term forecasting (1-3 step(s))", "medium-term forecasting (4-8 steps)", and "long-term forecasting (more than 8 steps) [63]". Here, "step" refers to the frequency of the time series, such as "daily", "weekly","monthly" or "quarterly". In this study, we perform 1,5 and 10-step-ahead tests to examine the short, middle and long term forecasting ability of the proposed model respectively. The chosen time window size for 1-day and 5-day-ahead forecasting is 5 and for 10-day-ahead forecasting is 10. The reason why we do so is that a typical week has five trading days and one year is about 250 trading days. We use 5 statistical measures, namely RMSE, MAE, MAPE, MSE, NMSE (see Appendix A.1), to evaluate forecasting performance. Figures 5.4 to 5.14, A.1 to A.11 and Tables 5.5 to 5.15 show the results of the experiments.

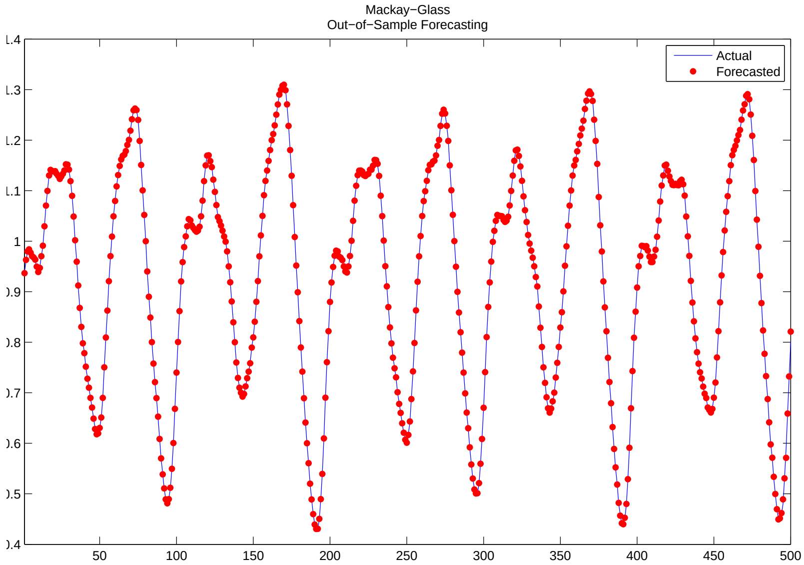



Figure 5.2: Forecasting result of Mackey-Glass Time Series (1-step-ahead)

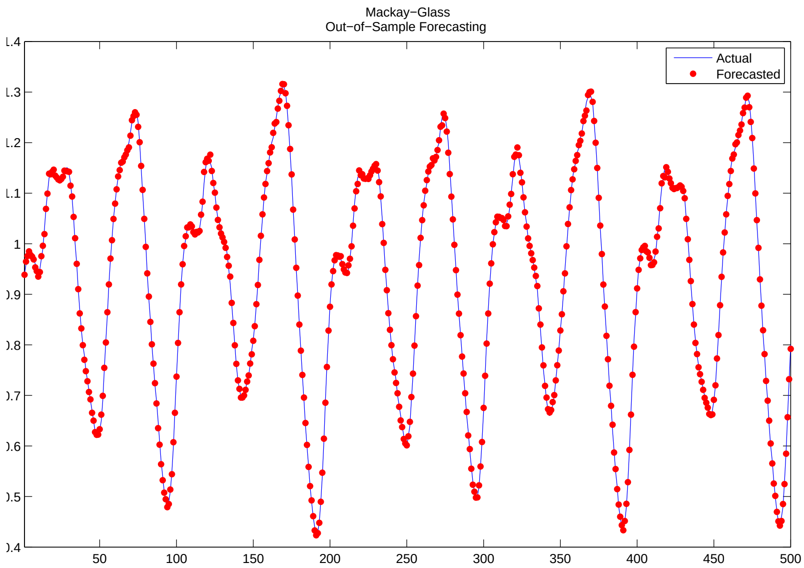



Figure 5.3: Forecasting result of Mackey-Glass Time Series (6-step-ahead)

Table 5.5: A Comparison of Forecasting Performance for Canadian Dollar/US Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0220 | 0.0202 | 1.7921 | 0.0005 | 1.3659 |
| NN+Wavelet Denoising | 0.0036 | 0.0027 | 0.2417 | 1.3e-5 | 0.0359 |
| NN+Wavelet Packet Denoising | 0.0236 | 0.0217 | 1.9290 | 0.0006 | 1.5705 |
| New Method(DWT) | 0.0004 | 0.0003 | 0.0261 | 1.3e-7 | 0.0004 |
| without statistical feature | 0.0031 | 0.0026 | 0.2273 | 9.7e-6 | 0.0214 |
| New Method(SWT)* | 0.0019 | 0.0015 | 0.1306 | 3.6e-6 | 0.0080 |
| 5-step-ahead | | | | | |
| NN | 0.0543 | 0.0433 | 3.8415 | 0.0029 | 8.2867 |
| NN+Wavelet Denoising | 0.0328 | 0.0308 | 2.7253 | 0.0011 | 3.0262 |
| NN+Wavelet Packet Denoising | 0.0666 | 0.0614 | 5.4395 | 0.0044 | 12.4938 |
| New Method(DWT) | 0.0020 | 0.0015 | 0.1336 | 4.0e-6 | 0.0113 |
| without statistical feature | 0.0144 | 0.0124 | 1.1014 | 0.0002 | 0.4579 |
| New Method(SWT)† | 0.0066 | 0.0053 | 0.4686 | 4.4e-5 | 0.0965 |
| 10-step-ahead | | | | | |
| NN | 0.0953 | 0.0928 | 8.2155 | 0.0091 | 25.5368 |
| NN+Wavelet Denoising | 0.0541 | 0.0426 | 3.7795 | 0.0029 | 8.2407 |
| NN+Wavelet Packet Denoising | 0.1066 | 0.1033 | 9.1422 | 0.0114 | 31.9762 |
| New Method(DWT) | 0.0059 | 0.0051 | 0.4538 | 3.4e-5 | 0.0964 |
| without statistical feature | 0.0216 | 0.0203 | 1.7950 | 0.0005 | 1.0295 |
| New Method(SWT)‡ | 0.0105 | 0.0090 | 0.8020 | 0.0001 | 0.2428 |

* The Kurtosis and Shannon Entropy features are used.
† The Skewness and Shannon Entropy features are used.
‡ The Mean feature is used.

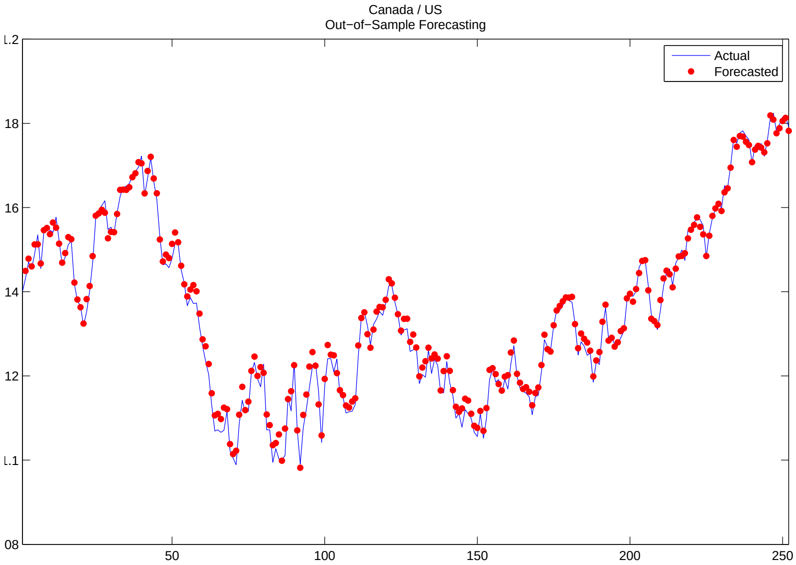



Figure 5.4: 1-step-ahead Forecasting Result for Canadian Dollar/US Dollar

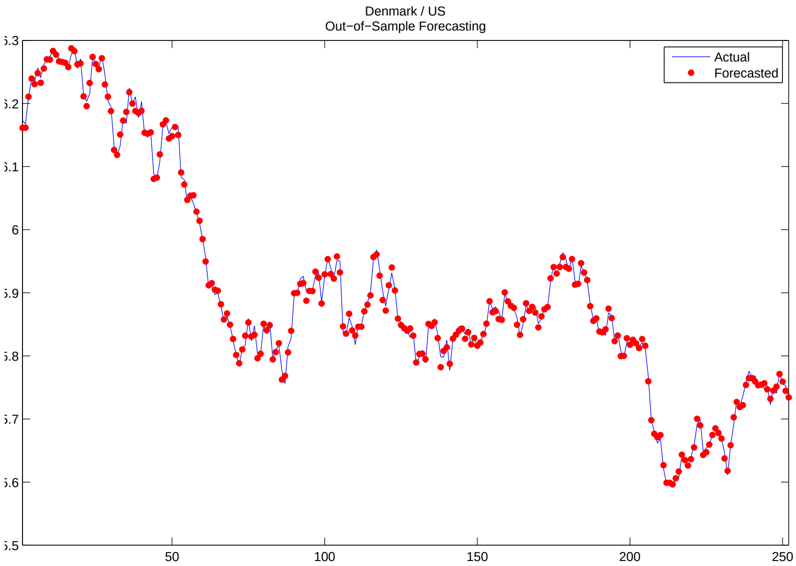



Figure 5.5: 1-step-ahead Forecasting Result for Danish Kroner/US Dollar

Table 5.6: A Comparison of Forecasting Performance for Danish Kroner/US Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0282 | 0.0215 | 0.3611 | 0.0008 | 0.0240 |
| NN+Wavelet Denoising | 0.0172 | 0.0138 | 0.2311 | 0.0003 | 0.0090 |
| NN+Wavelet Packet Denoising | 0.0172 | 0.0138 | 0.2311 | 0.0003 | 0.0090 |
| New Method(DWT) | 0.0010 | 0.0008 | 0.0139 | 1.1e-6 | 3.2e-5 |
| without statistical feature | 0.0082 | 0.0064 | 0.1076 | 6.9e-5 | 0.0021 |
| New Method(SWT)* | 0.0069 | 0.0054 | 0.0917 | 4.7e-5 | 0.0015 |
| 5-step-ahead | | | | | |
| NN | 0.0650 | 0.0519 | 0.8736 | 0.0042 | 0.1276 |
| NN+Wavelet Denoising | 0.0358 | 0.0280 | 0.4702 | 0.0013 | 0.0387 |
| NN+Wavelet Packet Denoising | 0.0369 | 0.0292 | 0.4902 | 0.0014 | 0.0411 |
| New Method(DWT) | 0.0110 | 0.0087 | 0.1465 | 0.0001 | 0.0036 |
| without statistical feature | 0.0303 | 0.0239 | 0.4037 | 0.0009 | 0.0288 |
| New Method(SWT)† | 0.0259 | 0.0206 | 0.3497 | 0.0007 | 0.0210 |
| 10-step-ahead | | | | | |
| NN | 0.0909 | 0.0739 | 1.2462 | 0.0083 | 0.2492 |
| NN+Wavelet Denoising | 0.0752 | 0.0600 | 1.0107 | 0.0057 | 0.1709 |
| NN+Wavelet Packet Denoising | 0.0874 | 0.0717 | 1.2128 | 0.0076 | 0.2309 |
| New Method(DWT) | 0.0159 | 0.0126 | 0.2122 | 0.0003 | 0.0076 |
| without statistical feature | 0.0391 | 0.0313 | 0.5301 | 0.0015 | 0.0480 |
| New Method(SWT)‡ | 0.0370 | 0.0298 | 0.5033 | 0.0014 | 0.0431 |

* The Mean Absolute Deviation and Mean features are used.

† The Kurtosis and Mean features are used.

‡ The Mean feature is used.

Table 5.7: A Comparison of Forecasting Performance for Japanese Yen/US Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.6238 | 0.4662 | 0.4015 | 0.3891 | 0.0895 |
| NN+Wavelet Denoising | 0.3589 | 0.2868 | 0.2469 | 0.1288 | 0.0296 |
| NN+Wavelet Packet Denoising | 0.3579 | 0.2863 | 0.2464 | 0.1281 | 0.0295 |
| New Method(DWT) | 0.0221 | 0.0173 | 0.0149 | 0.0005 | 0.0001 |
| without statistical feature | 0.1721 | 0.1374 | 0.1178 | 0.0296 | 0.0053 |
| New Method(SWT)* | 0.1489 | 0.1176 | 0.1009 | 0.0222 | 0.0040 |
| 5-step-ahead | | | | | |
| NN | 1.7802 | 1.4118 | 1.2204 | 3.1693 | 0.7293 |
| NN+Wavelet Denoising | 0.8094 | 0.6210 | 0.5350 | 0.6550 | 0.1507 |
| NN+Wavelet Packet Denoising | 0.7278 | 0.5708 | 0.4912 | 0.5297 | 0.1219 |
| New Method(DWT) | 0.2149 | 0.1706 | 0.1470 | 0.0462 | 0.0106 |
| without statistical feature | 0.7248 | 0.5925 | 0.5080 | 0.5253 | 0.0948 |
| New Method(SWT)† | 0.5452 | 0.4458 | 0.3819 | 0.2972 | 0.0536 |
| 10-step-ahead | | | | | |
| NN | 2.5473 | 2.0465 | 1.7722 | 6.4888 | 1.4931 |
| NN+Wavelet Denoising | 2.2621 | 1.8243 | 1.5796 | 5.1172 | 1.1775 |
| NN+Wavelet Packet Denoising | 2.1449 | 1.7570 | 1.5199 | 4.6007 | 1.0586 |
| New Method(DWT) | 0.3650 | 0.2814 | 0.2423 | 0.1333 | 0.0307 |
| without statistical feature | 0.9120 | 0.7310 | 0.6272 | 0.8317 | 0.1501 |
| New Method(SWT)‡ | 0.8191 | 0.6638 | 0.5697 | 0.6708 | 0.1211 |

* The Skewness and Shannon Entropy features are used.
† The Turning Points and Shannon Entropy features are used.
‡ The Mean Absolute Deviation and Shannon Entropy features are used.

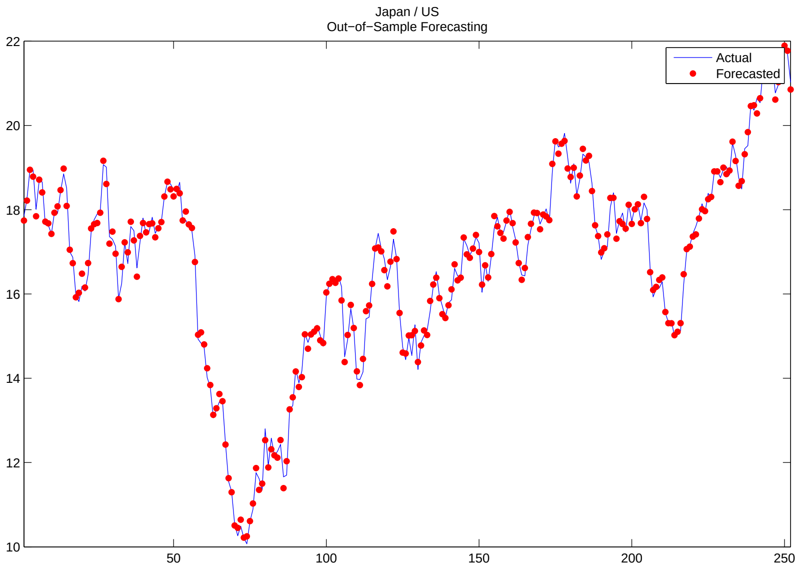



Figure 5.6: 1-step-ahead Forecasting Result for Japanese Yen/US Dollar

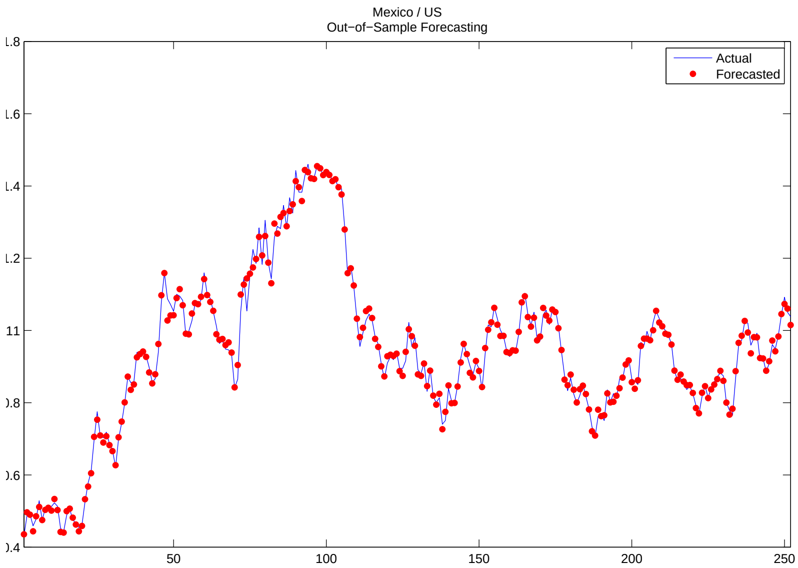



Figure 5.7: 1-step-ahead Forecasting Result for Mexican New Pesos/US Dollar

Table 5.8: A Comparison of Forecasting Performance for Mexican New Pesos/US Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0506 | 0.0402 | 0.3674 | 0.0026 | 0.0419 |
| NN+Wavelet Denoising | 0.0301 | 0.0242 | 0.2210 | 0.0009 | 0.0148 |
| NN+Wavelet Packet Denoising | 0.0301 | 0.0242 | 0.2211 | 0.0009 | 0.0148 |
| New Method(DWT) | 0.0019 | 0.0014 | 0.0128 | 3.6e-6 | 5.9e-5 |
| without statistical feature | 0.0480 | 0.0163 | 0.1485 | 0.0023 | 0.0464 |
| New Method(SWT)* | 0.0153 | 0.0110 | 0.1000 | 0.0002 | 0.0047 |
| 5-step-ahead | | | | | |
| NN | 0.1337 | 0.1085 | 0.9940 | 0.0179 | 0.2927 |
| NN+Wavelet Denoising | 0.0629 | 0.0498 | 0.4558 | 0.0040 | 0.0647 |
| NN+Wavelet Packet Denoising | 0.0599 | 0.0479 | 0.4386 | 0.0036 | 0.0587 |
| New Method(DWT) | 0.0203 | 0.0149 | 0.1358 | 0.0004 | 0.0068 |
| without statistical feature | 0.0606 | 0.0473 | 0.4324 | 0.0037 | 0.0737 |
| New Method(SWT)† | 0.0508 | 0.0397 | 0.3628 | 0.0026 | 0.0519 |
| 10-step-ahead | | | | | |
| NN | 0.2206 | 0.1672 | 1.5353 | 0.0487 | 0.7961 |
| NN+Wavelet Denoising | 0.1838 | 0.1433 | 1.3131 | 0.0338 | 0.5530 |
| NN+Wavelet Packet Denoising | 0.1713 | 0.1369 | 1.2565 | 0.0293 | 0.4801 |
| New Method(DWT) | 0.0303 | 0.0242 | 0.2205 | 0.0009 | 0.0151 |
| without statistical feature | 0.0823 | 0.0644 | 0.5878 | 0.0068 | 0.1362 |
| New Method(SWT)‡ | 0.0786 | 0.0629 | 0.5733 | 0.0062 | 0.1242 |

* The Variance and Mean features are used.
† The Kurtosis and Shannon Entropy features are used.
‡ The Shannon Entropy feature is used.

Table 5.9: A Comparison of Forecasting Performance for Norwegian Kroner/US Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0407 | 0.0321 | 0.4991 | 0.0017 | 0.0340 |
| NN+Wavelet Denoising | 0.0252 | 0.0203 | 0.3156 | 0.0006 | 0.0131 |
| NN+Wavelet Packet Denoising | 0.0252 | 0.0203 | 0.3158 | 0.0006 | 0.0131 |
| New Method(DWT) | 0.0013 | 0.0010 | 0.0162 | 1.7e-6 | 3.4e-5 |
| without statistical feature | 0.0116 | 0.0090 | 0.1416 | 0.0001 | 0.0030 |
| New Method(SWT)* | 0.0106 | 0.0084 | 0.1319 | 0.0001 | 0.0025 |
| 5-step-ahead | | | | | |
| NN | 0.1050 | 0.0874 | 1.3621 | 0.0110 | 0.2267 |
| NN+Wavelet Denoising | 0.0520 | 0.0423 | 0.6594 | 0.0027 | 0.0556 |
| NN+Wavelet Packet Denoising | 0.0587 | 0.0472 | 0.7353 | 0.0034 | 0.0710 |
| New Method(DWT) | 0.0147 | 0.0119 | 0.1860 | 0.0002 | 0.0045 |
| without statistical feature | 0.0528 | 0.0429 | 0.6714 | 0.0028 | 0.0627 |
| New Method(SWT)† | 0.0397 | 0.0323 | 0.5057 | 0.0016 | 0.0355 |
| 10-step-ahead | | | | | |
| NN | 0.1561 | 0.1267 | 1.9829 | 0.0244 | 0.5015 |
| NN+Wavelet Denoising | 0.1528 | 0.1203 | 1.8793 | 0.0234 | 0.4805 |
| NN+Wavelet Packet Denoising | 0.1395 | 0.1145 | 1.7863 | 0.0194 | 0.4001 |
| New Method(DWT) | 0.0260 | 0.0206 | 0.3214 | 0.0007 | 0.0139 |
| without statistical feature | 0.0615 | 0.0497 | 0.7786 | 0.0038 | 0.0851 |
| New Method(SWT)‡ | 0.0500 | 0.0403 | 0.6302 | 0.0025 | 0.0561 |

* The Variance and Shannon Entropy features are used.
† The Kurtosis and Shannon Entropy features are used.
‡ The Shannon Entropy feature is used.

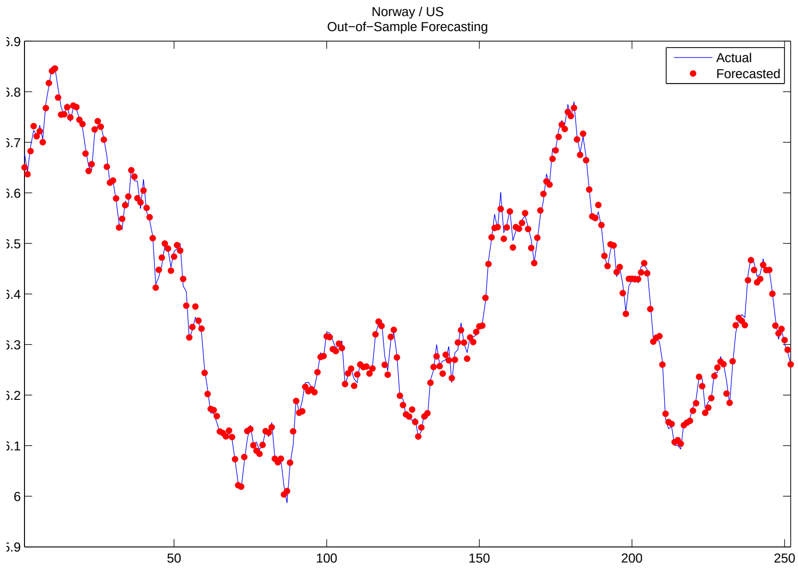



Figure 5.8: 1-step-ahead Forecasting Result for Norwegian Kroner/US Dollar

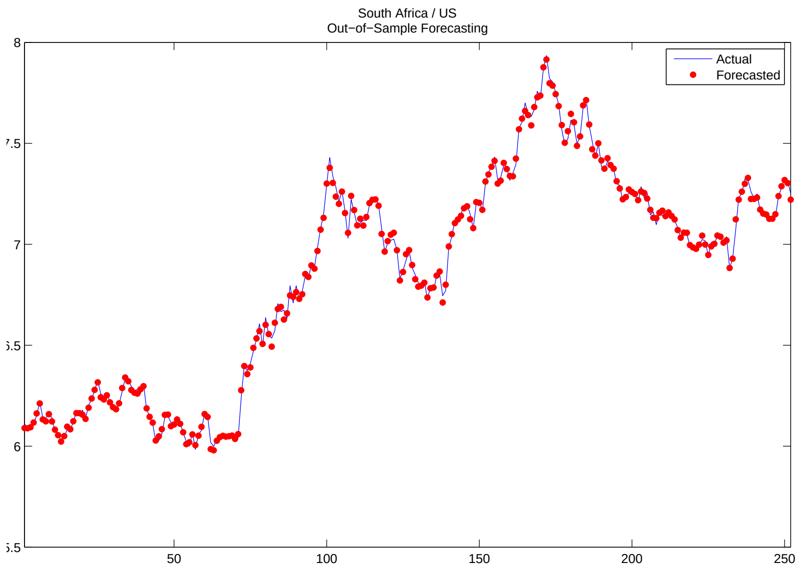



Figure 5.9: 1-step-ahead Forecasting Result for South African Rand/US Dollar

Table 5.10: A Comparison of Forecasting Performance for South African Rand/US Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0663 | 0.0516 | 0.7549 | 0.0044 | 0.0142 |
| NN+Wavelet Denoising | 0.0400 | 0.0313 | 0.4586 | 0.0016 | 0.0052 |
| NN+Wavelet Packet Denoising | 0.0400 | 0.0312 | 0.4577 | 0.0016 | 0.0051 |
| New Method(DWT) | 0.0027 | 0.0020 | 0.0292 | 7.2e-6 | 2.3e-5 |
| without statistical feature | 0.0216 | 0.0168 | 0.2438 | 0.0005 | 0.0017 |
| New Method(SWT)* | 0.0177 | 0.0140 | 0.2032 | 0.0003 | 0.0011 |
| 5-step-ahead | | | | | |
| NN | 0.1507 | 0.1149 | 1.6804 | 0.0227 | 0.0732 |
| NN+Wavelet Denoising | 0.0928 | 0.0706 | 1.0333 | 0.0086 | 0.0278 |
| NN+Wavelet Packet Denoising | 0.0895 | 0.0686 | 1.0028 | 0.0080 | 0.0258 |
| New Method(DWT) | 0.0272 | 0.0205 | 0.3007 | 0.0007 | 0.0024 |
| without statistical feature | 0.0718 | 0.0554 | 0.8041 | 0.0052 | 0.0185 |
| New Method(SWT)† | 0.0634 | 0.0485 | 0.7023 | 0.0040 | 0.0144 |
| 10-step-ahead | | | | | |
| NN | 0.2173 | 0.1748 | 2.5514 | 0.0472 | 0.1522 |
| NN+Wavelet Denoising | 0.1894 | 0.1495 | 2.1859 | 0.0359 | 0.1156 |
| NN+Wavelet Packet Denoising | 0.1876 | 0.1479 | 2.1608 | 0.0352 | 0.1134 |
| New Method(DWT) | 0.0416 | 0.0317 | 0.4673 | 0.0017 | 0.0056 |
| without statistical feature | 0.1077 | 0.0848 | 1.2341 | 0.0116 | 0.0416 |
| New Method(SWT)‡ | 0.0932 | 0.0740 | 1.0728 | 0.0087 | 0.0312 |

* The Mean and Historical Volatility features are used.

† The Kurtosis and Mean features are used.

‡ The Shannon Entropy and Historical Volatility features are used.

Table 5.11: A Comparison of Forecasting Performance for Swiss Francs/US Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0070 | 0.0054 | 0.4315 | 5.0e-5 | 0.0455 |
| NN+Wavelet Denoising | 0.0044 | 0.0035 | 0.2787 | 2.0e-5 | 0.0182 |
| NN+Wavelet Packet Denoising | 0.0070 | 0.0055 | 0.4351 | 4.9e-5 | 0.0452 |
| New Method(DWT) | 0.0002 | 0.0002 | 0.0142 | 5.4e-8 | 5.0e-5 |
| without statistical feature | 0.0020 | 0.0015 | 0.1236 | 3.9e-6 | 0.0038 |
| New Method(SWT)* | 0.0017 | 0.0013 | 0.1063 | 2.9e-6 | 0.0028 |
| 5-step-ahead | | | | | |
| NN | 0.0167 | 0.0134 | 1.0673 | 0.0003 | 0.2584 |
| NN+Wavelet Denoising | 0.0082 | 0.0064 | 0.5093 | 6.8e-5 | 0.0625 |
| NN+Wavelet Packet Denoising | 0.0158 | 0.0125 | 0.9996 | 0.0003 | 0.2323 |
| New Method(DWT) | 0.0023 | 0.0019 | 0.1486 | 5.4e-6 | 0.0050 |
| without statistical feature | 0.0080 | 0.0065 | 0.5183 | 6.4e-5 | 0.0617 |
| New Method(SWT)† | 0.0064 | 0.0050 | 0.3994 | 4.0e-5 | 0.0391 |
| 10-step-ahead | | | | | |
| NN | 0.0241 | 0.0195 | 1.5593 | 0.0006 | 0.5383 |
| NN+Wavelet Denoising | 0.0201 | 0.0162 | 1.2980 | 0.0004 | 0.3743 |
| NN+Wavelet Packet Denoising | 0.0216 | 0.0172 | 1.3781 | 0.0005 | 0.4313 |
| New Method(DWT) | 0.0039 | 0.0031 | 0.2450 | 1.5e-5 | 0.0140 |
| without statistical feature | 0.0086 | 0.0071 | 0.5632 | 7.5e-5 | 0.0724 |
| New Method(SWT)‡ | 0.0078 | 0.0062 | 0.4989 | 6.0e-5 | 0.0584 |

* The Mean Absolute Deviation and Shannon Entropy features are used.

† The Turning Points and Shannon Entropy features are used.

‡ The Skewness and Shannon Entropy features are used.

We can see, from Tables 5.5 to 5.15, that the proposed model can provide accurate short term forecast. However, the results for medium-term horizon are less accurate and for long term horizon are much less accurate. This is because foreign currency market is a complex system [70] that is affected by many factors. When the time horizon increases, the forecast involves the actions and interactions of more and more variables, making the task becomes exponentially more difficult and finally impossible. Furthermore, we notice that the performance of the wavelet multi-resolution based method consistently towers above the pure neural network model and the wavelet-denoising based approaches. This is contrary to the results that we got using Mackey-Glass time series (see Section 5.1). We consider that this is due to two reasons: firstly, the forecasting ability of our method is seriously limited by the experimental set-up of the 6-step-ahead forecasting. In the proposed model, we mainly use two approaches to achieve

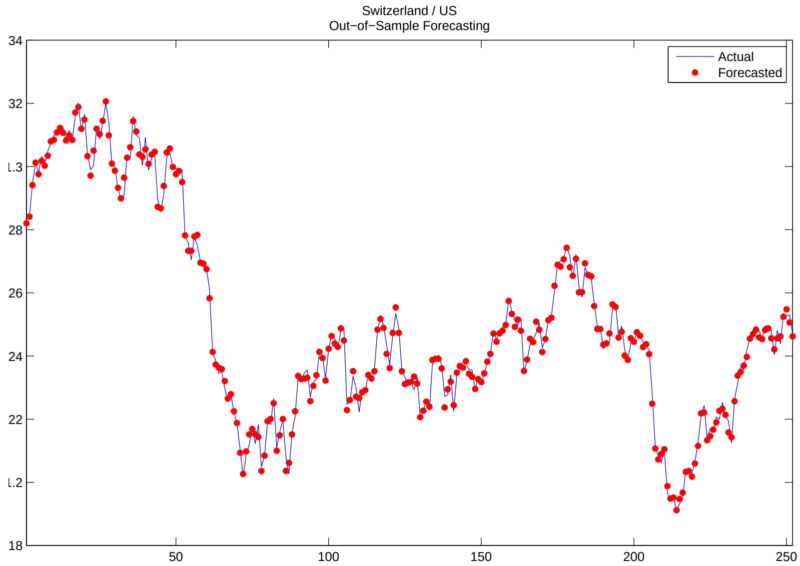



Figure 5.10: 1-step-ahead Forecasting Result for Swiss Francs/US Dollar

high-accuracy forecast: wavelet multi-resolution analysis and additional statistical features. However, according to the experimental set-up, we can use only 4 discrete values i.e. $x(t)$, $x(t-6)$, $x(t-12)$, $x(t-18)$ as the inputs of the neural networks to predict the value of $x(t+6)$, this makes both methods meaningless! To apply wavelet analysis and features extraction to 4 discrete values is a waste of time! Table 5.3 indicates that the performance of our system is slightly worse than back-propagation neural network, but this is definitely not the real performance of our model. Secondly, we think that Mackey-Glass time series is not a reliable benchmark for this study. Because real exchange rate time series is characterized with high volatility while Mackey-Glass time series changes direction slowly. Therefore, a method that gives good results for Mackey-Glass time series does not necessarily provide the same performance for real-world data sets. Figure 5.15 shows a comparison of the historical volatility of Mackey-Glass time series and a real exchange rate time series (Canadian Dollar/US Dollar). We can easily see the difference: Mackey-Glass volatility curve has some regular patterns, whereas the exchange rate volatility curve does not have any regular pattern. We believe that this is why Mackey-Glass time series is used as a benchmark in the areas of neural networks, fuzzy systems and hybrid systems but not in the fields of economics and finance. As the Mackey-Glass time series prediction task cannot fully reflect the forecasting ability of the proposed model, it is necessary to perform comparative experiments to assess the relative performance of our method.

Table 5.12: A Comparison of Forecasting Performance for US Dollar/Australian Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0040 | 0.0032 | 0.4247 | 1.6e-5 | 0.0530 |
| NN+Wavelet Denoising | 0.0024 | 0.0019 | 0.2510 | 5.5e-6 | 0.0181 |
| NN+Wavelet Packet Denoising | 0.0024 | 0.0019 | 0.2509 | 5.5e-6 | 0.0181 |
| New Method(DWT) | 0.0001 | 0.0001 | 0.0157 | 2.2e-8 | 7.2e-5 |
| without statistical feature | 0.0011 | 0.0009 | 0.1138 | 1.3e-6 | 0.0034 |
| New Method(SWT)* | 0.0010 | 0.0008 | 0.1050 | 1.0e-6 | 0.0027 |
| 5-step-ahead | | | | | |
| NN | 0.0111 | 0.0091 | 1.2164 | 0.0001 | 0.4044 |
| NN+Wavelet Denoising | 0.0053 | 0.0042 | 0.5587 | 2.9e-5 | 0.0934 |
| NN+Wavelet Packet Denoising | 0.0053 | 0.0042 | 0.5611 | 2.8e-5 | 0.0925 |
| New Method(DWT) | 0.0016 | 0.0012 | 0.1661 | 2.5e-6 | 0.0083 |
| without statistical feature | 0.0067 | 0.0050 | 0.6520 | 4.5e-5 | 0.1211 |
| New Method(SWT)† | 0.0050 | 0.0039 | 0.5062 | 2.5e-5 | 0.0682 |
| 10-step-ahead | | | | | |
| NN | 0.0206 | 0.0168 | 2.2476 | 0.0004 | 1.3856 |
| NN+Wavelet Denoising | 0.0127 | 0.0106 | 1.4094 | 0.0002 | 0.5308 |
| NN+Wavelet Packet Denoising | 0.0134 | 0.0111 | 1.4866 | 0.0002 | 0.5853 |
| New Method(DWT) | 0.0025 | 0.0019 | 0.2585 | 6.1e-6 | 0.0199 |
| without statistical feature | 0.0089 | 0.0067 | 0.8761 | 8.0e-5 | 0.2159 |
| New Method(SWT)‡ | 0.0069 | 0.0053 | 0.6892 | 4.7e-5 | 0.1279 |

* The Turning Points and Mean features are used.
† The Shannon Entropy feature is used.
‡ The Mean feature is used.

## 5.3 Comparison With Other Research

There are a large number of publications dealing with issue of exchange rate forecasting. However, different publications use different data sets and experimental set-ups. To make the comparison with others' work fair, we need to test our system with the same data set and experimental setting. This is not a easy thing to do, because some data sets are not available for downloading and many publications, in fact, do not contain complete information about the experimental set-up (see [63] Tables 2,3,4). Therefore, we selected 9 publications that are suitable for this study. In the following sections, we will present the results of the comparison in detail.

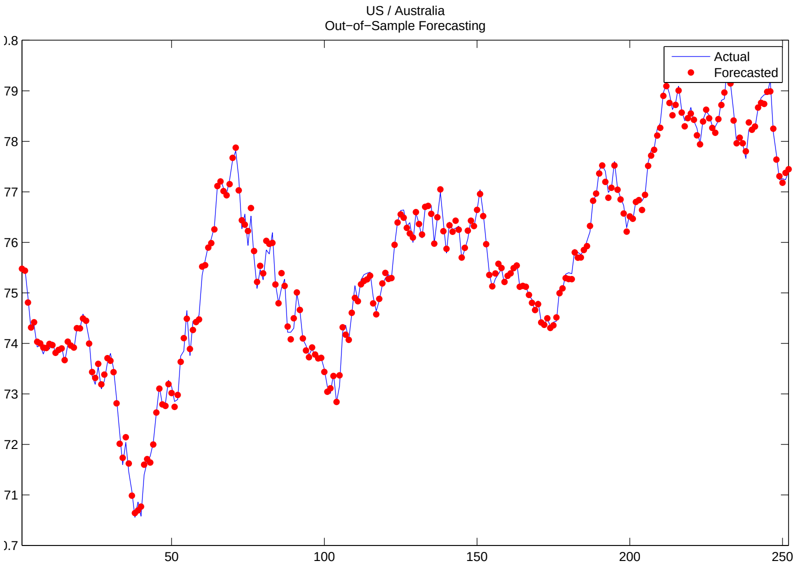



Figure 5.11: 1-step-ahead Forecasting Result for US Dollar/Australian Dollar

### Experiment One

In this experiment, we compared the performance of our model with the approaches from [14]. Below is a short description of [14] and the results are shown in Tables 5.17,5.18 and Figures 5.16,5.17.

[14] presents a new model that combines economic theory and machinery learning to predict exchange rates. The model employed a genetic algorithm that is constrained by structural exchange rate models to optimize combinations of input variables of neural network predictor. The experimental results show that the new method can provide more accurate forecast than traditional machine learning and Naive Forecast models.

Table 5.13: A Comparison of Forecasting Performance for US Dollar/Euro

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0059 | 0.0046 | 0.3679 | 3.6e-5 | 0.0249 |
| NN+Wavelet Denoising | 0.0036 | 0.0029 | 0.2305 | 1.3e-5 | 0.0092 |
| NN+Wavelet Packet Denoising | 0.0026 | 0.0021 | 0.1647 | 6.7e-6 | 0.0047 |
| New Method(DWT) | 0.0002 | 0.0002 | 0.0140 | 5.5e-8 | 3.8e-5 |
| without statistical feature | 0.0017 | 0.0013 | 0.1013 | 2.8e-6 | 0.0020 |
| New Method(SWT)* | 0.0015 | 0.0012 | 0.0928 | 2.3e-6 | 0.0017 |
| 5-step-ahead | | | | | |
| NN | 0.0145 | 0.0119 | 0.9496 | 0.0002 | 0.1480 |
| NN+Wavelet Denoising | 0.0074 | 0.0058 | 0.4626 | 5.5e-5 | 0.0383 |
| NN+Wavelet Packet Denoising | 0.0130 | 0.0105 | 0.8314 | 0.0002 | 0.1184 |
| New Method(DWT) | 0.0023 | 0.0018 | 0.1425 | 5.4e-6 | 0.0038 |
| without statistical feature | 0.0066 | 0.0052 | 0.4074 | 4.3e-5 | 0.0315 |
| New Method(SWT)† | 0.0056 | 0.0045 | 0.3550 | 3.1e-5 | 0.0228 |
| 10-step-ahead | | | | | |
| NN | 0.0206 | 0.0166 | 1.3199 | 0.0004 | 0.2980 |
| NN+Wavelet Denoising | 0.0177 | 0.0146 | 1.1543 | 0.0003 | 0.2194 |
| NN+Wavelet Packet Denoising | 0.0206 | 0.0170 | 1.3465 | 0.0004 | 0.2969 |
| New Method(DWT) | 0.0035 | 0.0028 | 0.2215 | 1.2e-5 | 0.0086 |
| without statistical feature | 0.0087 | 0.0069 | 0.5483 | 7.5e-5 | 0.0546 |
| New Method(SWT)‡ | 0.0078 | 0.0062 | 0.4944 | 6.1e-5 | 0.0444 |

* The Skewness and Mean features are used.

† The Mean feature is used.

‡ The Skewness and Shannon Entropy features are used.

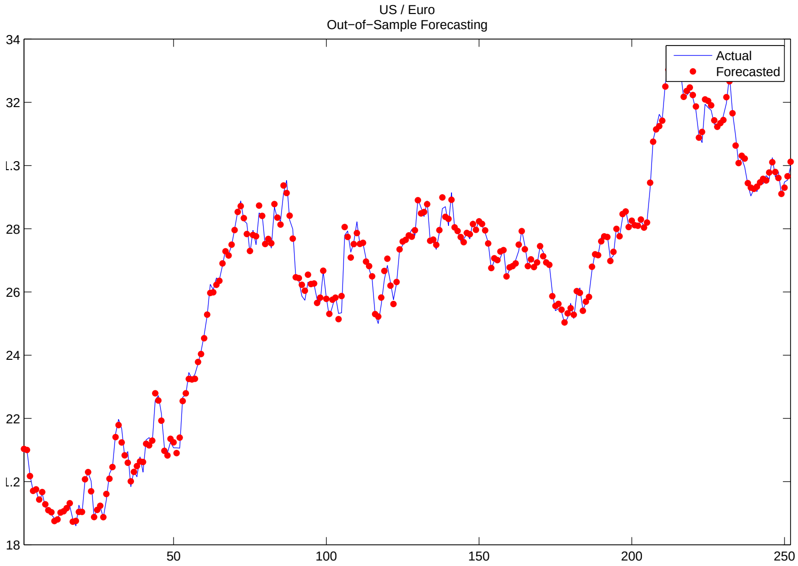



Figure 5.12: 1-step-ahead Forecasting Result for US Dollar/Euro

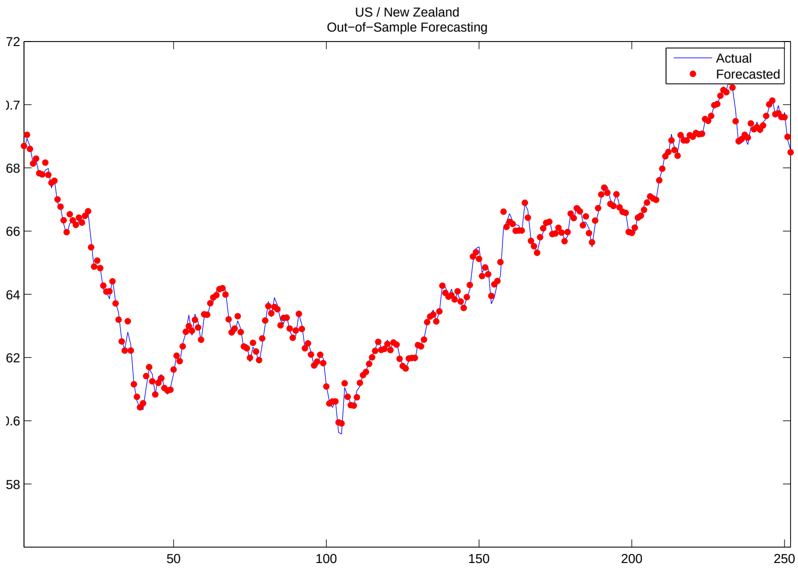



Figure 5.13: 1-step-ahead Forecasting Result for US Dollar/New Zealand Dollar

Table 5.14: A Comparison of Forecasting Performance for US Dollar/New Zealand Dollar

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0045 | 0.0035 | 0.5436 | 2.0e-5 | 0.0267 |
| NN+Wavelet Denoising | 0.0027 | 0.0022 | 0.3382 | 7.3e-6 | 0.0098 |
| NN+Wavelet Packet Denoising | 0.0027 | 0.0022 | 0.3381 | 7.3e-6 | 0.0098 |
| New Method(DWT) | 0.0002 | 0.0001 | 0.0220 | 3.4e-8 | 4.6e-5 |
| without statistical feature | 0.0015 | 0.0012 | 0.1836 | 2.4e-6 | 0.0030 |
| New Method(SWT)* | 0.0013 | 0.0010 | 0.1592 | 1.8e-6 | 0.0022 |
| 5-step-ahead | | | | | |
| NN | 0.0131 | 0.0104 | 1.6120 | 0.0002 | 0.2303 |
| NN+Wavelet Denoising | 0.0066 | 0.0053 | 0.8258 | 4.4e-5 | 0.0585 |
| NN+Wavelet Packet Denoising | 0.0065 | 0.0052 | 0.8168 | 4.3e-5 | 0.0572 |
| New Method(DWT) | 0.0017 | 0.0013 | 0.2033 | 2.9e-6 | 0.0039 |
| without statistical feature | 0.0053 | 0.0039 | 0.6037 | 2.8e-5 | 0.0342 |
| New Method(SWT)† | 0.0044 | 0.0033 | 0.5125 | 1.9e-5 | 0.0237 |
| 10-step-ahead | | | | | |
| NN | 0.0196 | 0.0153 | 2.3943 | 0.0004 | 0.5120 |
| NN+Wavelet Denoising | 0.0141 | 0.0114 | 1.7674 | 0.0002 | 0.2657 |
| NN+Wavelet Packet Denoising | 0.0171 | 0.0140 | 2.1800 | 0.0003 | 0.3917 |
| New Method(DWT) | 0.0028 | 0.0022 | 0.3374 | 7.8e-6 | 0.0105 |
| without statistical feature | 0.0065 | 0.0051 | 0.7858 | 4.3e-5 | 0.0531 |
| New Method(SWT)‡ | 0.0058 | 0.0045 | 0.6996 | 3.4e-5 | 0.0421 |

* The Skewness and Mean features are used.

† The Turning Points and Mean features are used.

‡ The Skewness and Shannon Entropy features are used.

Table 5.15: A Comparison of Forecasting Performance for US Dollar/British Pound

| Method | RMSE | MAE | MAPE | MSE | NMSE |
|---|---|---|---|---|---|
| 1-step-ahead | | | | | |
| NN | 0.0094 | 0.0072 | 0.3914 | 8.9e-5 | 0.0187 |
| NN+Wavelet Denoising | 0.0055 | 0.0043 | 0.2351 | 3.0e-5 | 0.0062 |
| NN+Wavelet Packet Denoising | 0.0094 | 0.0072 | 0.3924 | 8.9e-5 | 0.0187 |
| New Method(DWT) | 0.0004 | 0.0003 | 0.0153 | 1.3e-7 | 2.8e-5 |
| without statistical feature | 0.0029 | 0.0023 | 0.1221 | 8.6e-6 | 0.0016 |
| New Method(SWT)* | 0.0025 | 0.0019 | 0.1037 | 6.1e-6 | 0.0012 |
| 5-step-ahead | | | | | |
| NN | 0.0582 | 0.0450 | 2.4936 | 0.0034 | 0.7100 |
| NN+Wavelet Denoising | 0.0123 | 0.0098 | 0.5268 | 0.0002 | 0.0318 |
| NN+Wavelet Packet Denoising | 0.0323 | 0.0266 | 1.4641 | 0.0010 | 0.2190 |
| New Method(DWT) | 0.0035 | 0.0028 | 0.1546 | 1.2e-5 | 0.0026 |
| without statistical feature | 0.0141 | 0.0111 | 0.5963 | 0.0002 | 0.0384 |
| New Method(SWT)† | 0.0132 | 0.0103 | 0.5568 | 0.0002 | 0.0336 |
| 10-step-ahead | | | | | |
| NN | 0.0456 | 0.0376 | 2.0535 | 0.0021 | 0.4360 |
| NN+Wavelet Denoising | 0.0506 | 0.0407 | 2.2503 | 0.0026 | 0.5350 |
| NN+Wavelet Packet Denoising | 0.0475 | 0.0386 | 2.1132 | 0.0023 | 0.4715 |
| New Method(DWT) | 0.0061 | 0.0050 | 0.2733 | 3.8e-5 | 0.0079 |
| without statistical feature | 0.0188 | 0.0138 | 0.7444 | 0.0004 | 0.0679 |
| New Method(SWT)‡ | 0.0150 | 0.0122 | 0.6572 | 0.0002 | 0.0434 |

* The Mean Absolute Deviation and Mean features are used.

† The Mean and Historical Volatility features are used.

‡ The Skewness and Shannon Entropy features are used.

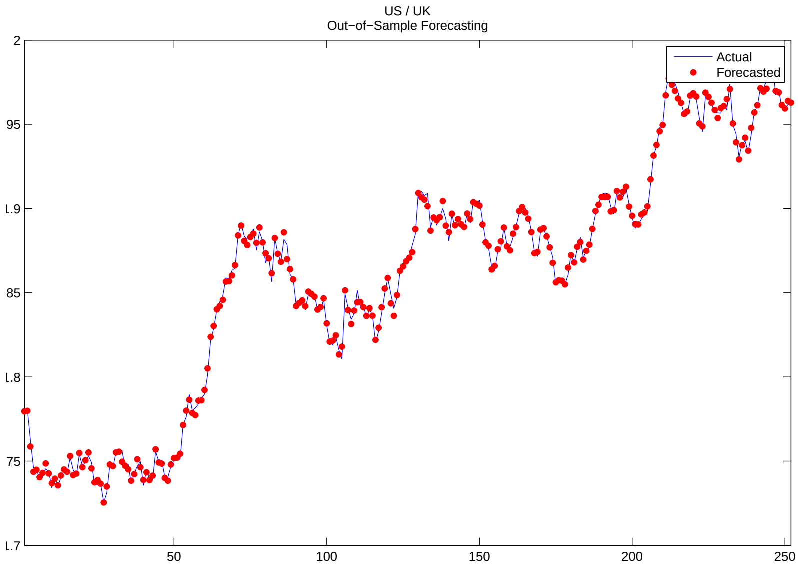
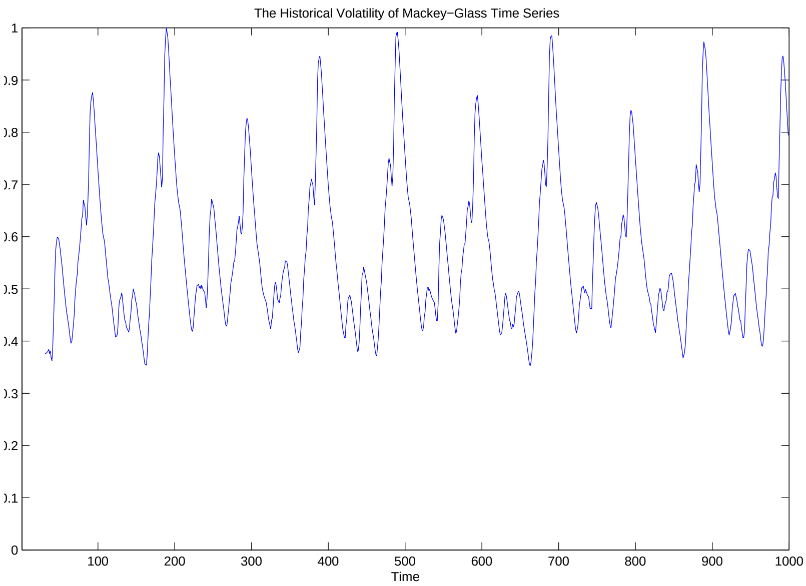
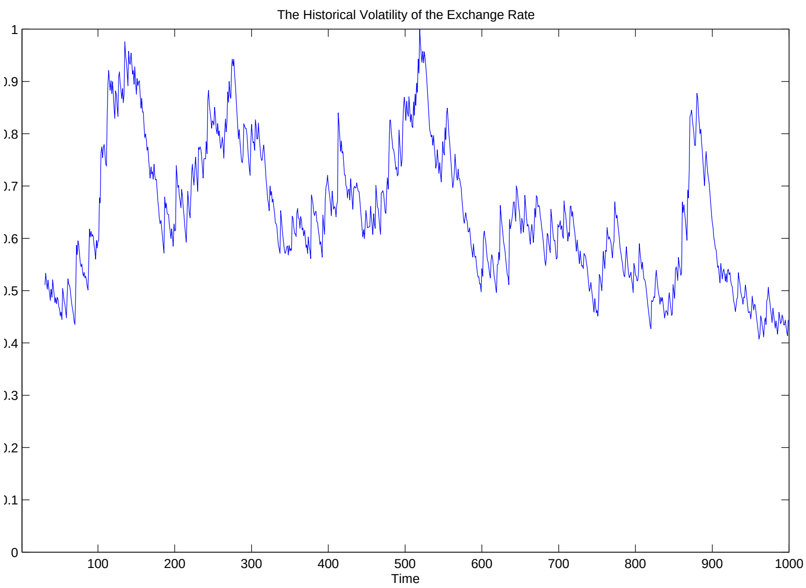



Figure 5.14: 1-step-ahead Forecasting Result for US Dollar/British Pound




(a) Mackey-Glass Time Series




(b) Exchange Rate Time Series (Canada/US)

Figure 5.15: Comparison of Historical Volatility

Table 5.16: Experiment settings one

| | |
|---|---|
| Exchange Rate(s) | Japan/US US/UK |
| Forecasting Horizon | One-Step-Ahead |
| Data Frequency | Monthly |
| Time Span | Mar-31-1991 to Sep-28-2001 |
| Training and Validation Sets Size | 98 |
| Testing Set Size | 30 |
| Performance Measure(s) | MAE MSE MAPE |

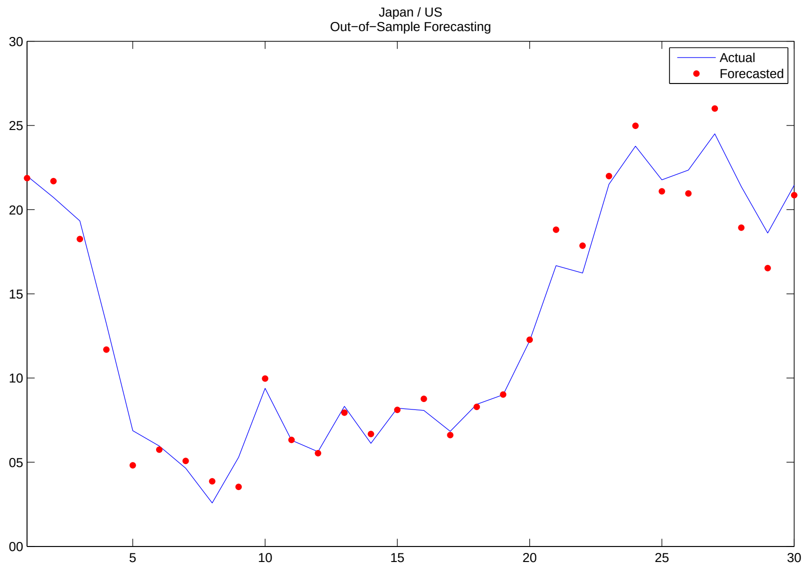



Figure 5.16: Japanese Yen/US Dollar

### Experiment Two

In this experiment, we compared the performance of our system with the methods from [80]. Below is a short description of [80] and the results are shown in Table 5.20 and Figure 5.18.

[80] proposes a Hybrid Support Vector Machine (HSVMG) forecasting model that consist of linear SVMs and nonlinear SVMs. In addition, the authors use genetic algorithms to select the parameters of the model. The experimental results are compared with the results from [66] to show the superiority of HSVMG model, and the comparative results indicate that the forecasting performance of the proposed model are better than Neural Networks (NN),Vector-valued Local Linear Approximation (VLLR) and Random Walk (RW) models.

Table 5.17: Forecasting performance for Japanese Yen/US Dollar

| Method | MAE | MSE | MAPE |
|---|---|---|---|
| Machine Learning* | 2.8760 | 13.0058 | 2.5127 |
| Combined Approach* | 2.7653 | 11.7405 | 2.4237 |
| Naive Forecast* | 3.0297 | 14.4797 | 2.6790 |
| NN+Wavelet Denoising | 2.6067 | 11.4316 | 2.3102 |
| NN+Wavelet Packet Denoising | 1.7244 | 6.5233 | 1.4938 |
| New Method(DWT) | 0.3464 | 0.2871 | 0.3008 |
| without statistical feature | 1.0352 | 2.7041 | 0.8985 |
| New Method(SWT)† | 0.8825 | 1.3082 | 0.7684 |

* The values in this row are taken from reference [14].
† The Variance feature is used.

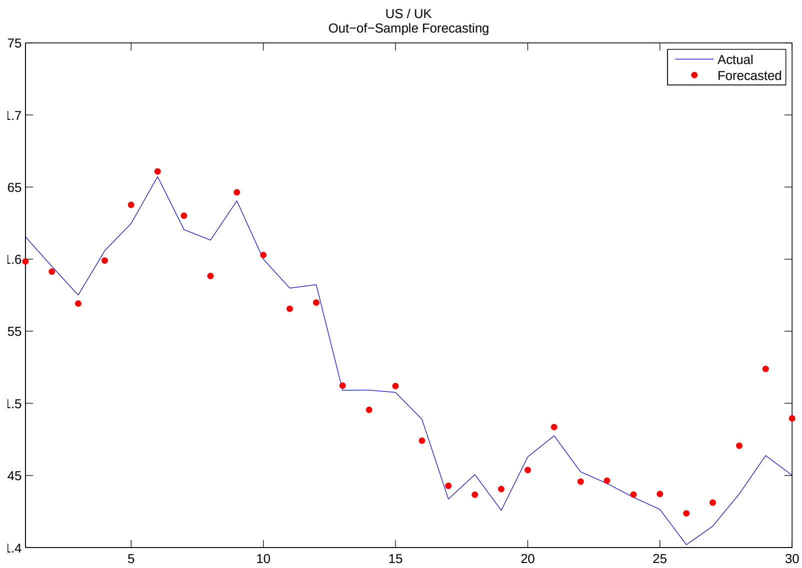



Figure 5.17: US Dollar/British Pound

**Experiment Three**

In this experiment, we compared the performance of our method with the results from [20]. Below is a short description of [20] and the results are shown in Table 5.22 and Figures 5.19,5.20,5.21.

[20] introduces Error Correction Neural Network (ECNN) models to forecast exchange rates. ECNN is a two-stage forecasting procedure which combines an econometric model and General Regression Neural Network (GRNN) in a sequential manner.In the first stage, a multivariate econometric model is used

Table 5.18: Forecasting performance for US Dollar/British Pound

| Method | MAE | MSE | MAPE |
|---|---|---|---|
| Machinery Learning* | 0.0248 | 0.0009 | 1.6376 |
| Combined Approach* | 0.0226 | 0.0008 | 1.4868 |
| Naive Forecast* | 0.0252 | 0.0009 | 1.6603 |
| NN+Wavelet Denoising | 0.0135 | 0.0003 | 0.8908 |
| NN+Wavelet Packet Denoising | 0.0373 | 0.0021 | 2.4805 |
| New Method(DWT) | 0.0018 | 4.6e-6 | 0.1186 |
| New Method(SWT) | 0.0134 | 0.0003 | 0.8963 |

* The values in this row are taken from reference [14].

Table 5.19: Experiment settings two

| | |
|---|---|
| Exchange Rate(s) | US/UK |
| Forecasting Horizon | One-Step-Ahead |
| Data Frequency | Monthly |
| Time Span | Jan-1973 to Oct-1995 |
| Training Set Size | 175 |
| Validation Set Size | 50 |
| Testing Set Size | 49 |
| Performance Measure(s) | NMSE |

to generate forecasts of the exchange rates and then,in the second stage, a GRNN is employed to correct the errors of the forecasts. In this research, the adopted econometric models are Multivariate Transfer Function (MTF), Generalized Method Of Moments (GMM) and Bayesian Vector Autoregression (BVAR) models. Results from the experiment demonstrate that the ECNN models generally outperform the single-stage econometric,GRNN and Random Walk models although the improvement obtained from using the two-stage models varies from currency to currency.

**Experiment Four**

In this experiment, we compared the performance of our model with the results from [69]. Below is a short description of [69] and the results are shown in Tables 5.24,5.25,5.26 and Figures 5.22,5.23,5.24.

[69] takes the natural logarithm of all variables in the data set and then uses the General Regression Neural Network (GRNN) to forecast the exchange rates. (In economics and finance, the introduction of the log-difference series is a com-

Table 5.20: Forecasting performance for US Dollar/British Pound

| Method | NMSE |
|---|---|
| HSVMG* | 0.4189 |
| NN* | 0.658 |
| VLLR* | 0.792 |
| RW* | 0.983 |
| NN+Wavelet Denoising | 0.0580 |
| NN+Wavelet Packet Denoising | 0.1469 |
| New Method(DWT) | 0.0007 |
| without statistical feature | 0.1118 |
| New Method(SWT)† | 0.0814 |

* The values in this row are taken from reference [80].
† The Turning Points feature is used.

Table 5.21: Experiment settings three

| | |
|---|---|
| Exchange Rate(s) | Canada/US Japan/US US/UK |
| Forecasting Horizon | One-Step-Ahead |
| Data Frequency | Monthly |
| Time Span | Jan-1980 to Dec-2001 |
| Training Set Size | 144 |
| Validation Set Size | 60 |
| Testing Set Size | 60 |
| Performance Measure(s) | RMSE |

mon method for overcoming non-stationarity as a prediction of a non-stationary series may be biased [11] [47]) The experimental results show that for all currencies that are tested in this study, GRNN models provide better forecasts than Multi-layered Feed-forward Neural Network (MLFN), Multivariate Transfer Function (MTF), and Random Walk models.

**Experiment Five**

In this experiment, we compared the performance of our model with the results from [62]. Below is a short description of [62] and the results are shown in Table 5.28 and Figures 5.25 and 5.26.

[62] presents a Nonlinear Ensemble (NE) forecasting model which integrates Generalized Linear Auto-Regression (GLAR) with Artificial Neural Network (ANN) to predict foreign exchange rates. Furthermore, the NE model's per-

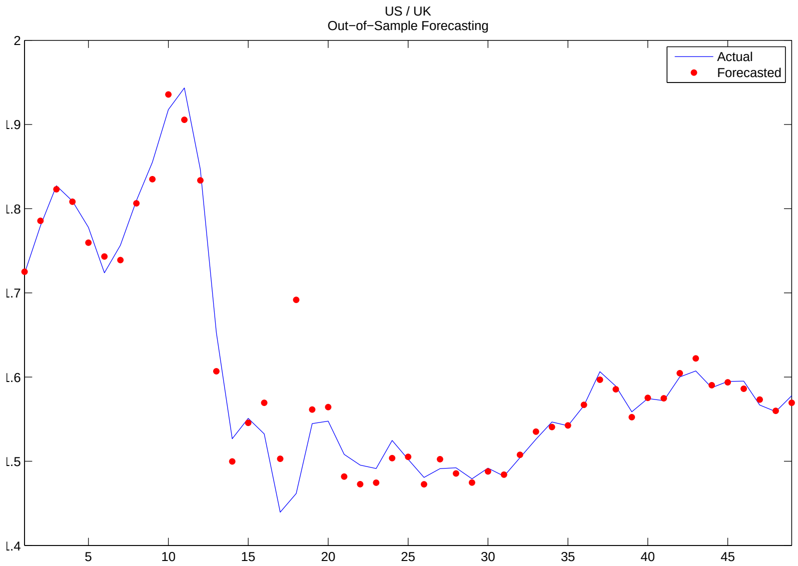



Figure 5.18: US Dollar/British Pound

formance is compared with GLAR, ANN, GLAR-ANN Hybrid, Equal Weights (EW) and Minimum-Error (ME) models. Empirical results reveal that the NE model produces better forecasting results than other methods.

Table 5.22: Forecasting performance for Canadian Dollar/US Dollar, Japanese Yen/US Dollar and US Dollar/British Pound

| Method | RMSE | | |
|---|---|---|---|
| | Canada/US | Japan/US | US/UK |
| MTF* | 0.0365 | 5.8294 | 0.0397 |
| BVAR* | 0.0329 | 5.6519 | 0.0372 |
| GMM* | 0.0352 | 5.9847 | 0.0359 |
| GRNN* | 0.0338 | 5.5240 | 0.0367 |
| ECNN-MTF* | 0.0334 | 5.4133 | 0.0355 |
| ECNN-BVAR* | 0.0291 | 5.0645 | 0.0348 |
| ECNN-GMM* | 0.0326 | 5.4648 | 0.0338 |
| Random Walk* | 0.0396 | 6.1830 | 0.0408 |
| NN+Wavelet Denoising | 0.0114 | 3.3492 | 0.0177 |
| NN+Wavelet Packet Denoising | 0.0089 | 4.3423 | 0.0376 |
| New Method(DWT) | 0.0080 | 0.5182 | 0.0017 |
| without statistical feature | - | 1.1507 | 0.0098 |
| New Method(SWT) | 0.0137 | 0.9967† | 0.0094‡ |

* The values in this row are taken from reference [20].
† The Variance and Shannon Entropy features are used.
‡ The Variance and Historical Volatility features are used.

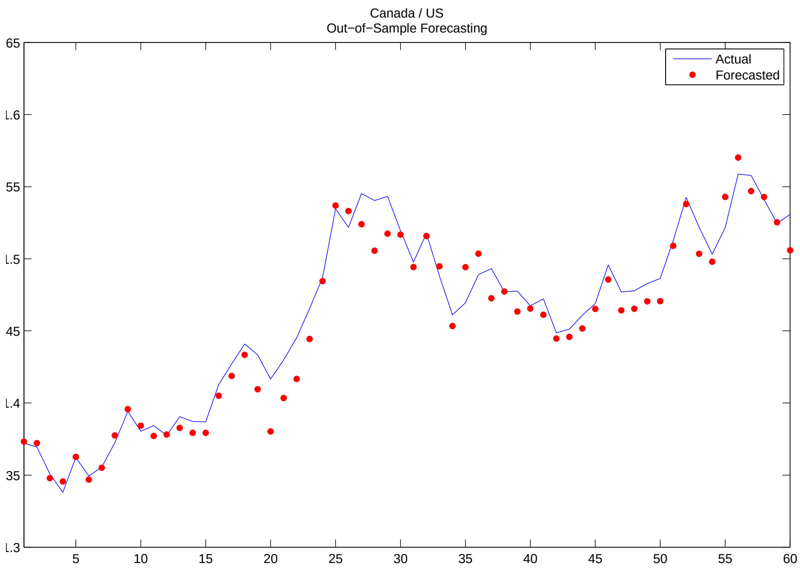



Figure 5.19: Canadian Dollar/US Dollar

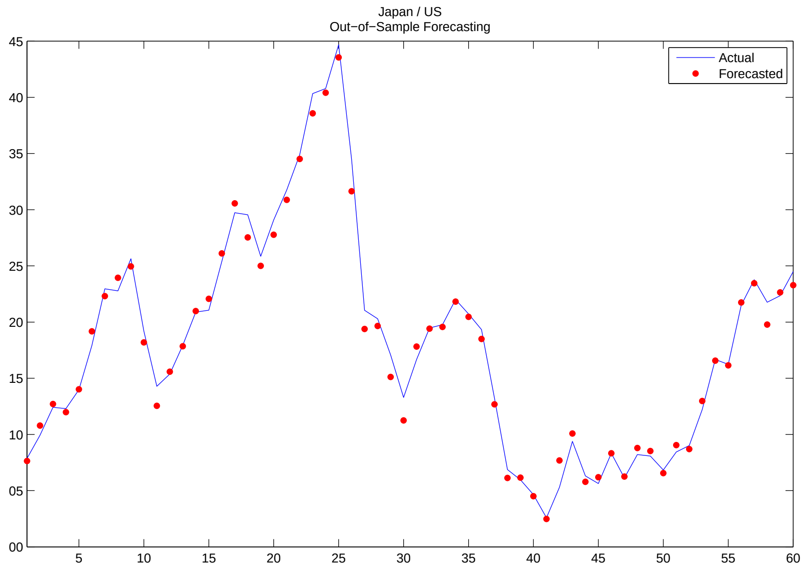



Figure 5.20: Japanese Yen/US Dollar

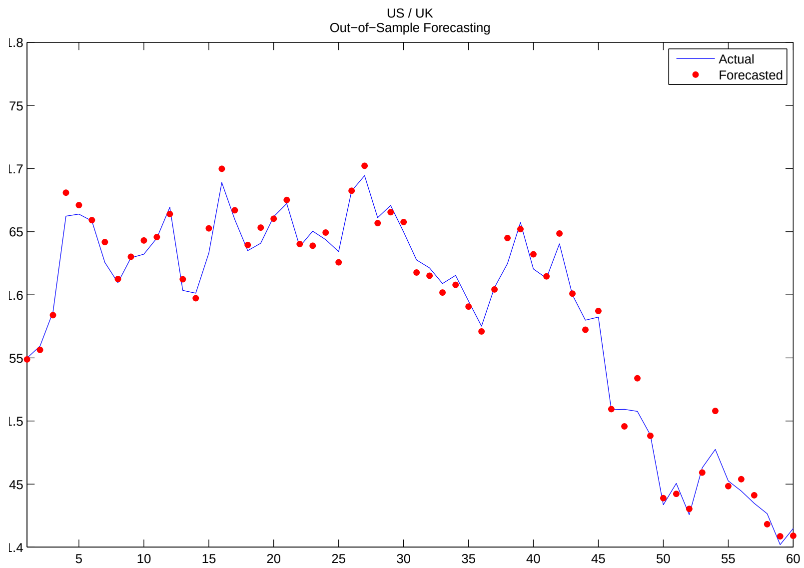



Figure 5.21: US Dollar/British Pound

Table 5.23: Experiment settings four

| | |
|---|---|
| Exchange Rate(s) | Canada/US* Japan/US* US/UK* |
| Forecasting Horizon | One-Step-Ahead |
| Data Frequency | Monthly |
| Time Span | Jan-1974 to Jul-1995 |
| Training Set Size | 129 |
| Validation Set Size | 65 |
| Testing Set Size | 65 |
| Performance Measure(s) | MAE RMSE |

* The natural logarithm of the exchange rates time series.

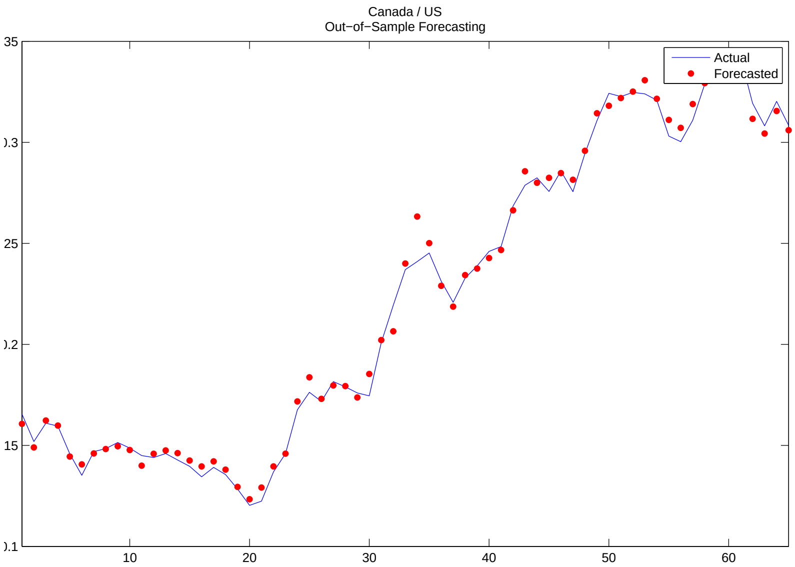



Figure 5.22: Canadian Dollar/US Dollar

Table 5.24: Forecasting performance for Canadian Dollar/US Dollar

| Method | MAE | RMSE |
|---|---|---|
| GRNN AR(1)* | 0.00866 | 0.01065 |
| GRNN AR(12)* | 0.00870 | 0.01056 |
| GRNN AR(1)-MUIP(1)* | 0.00868 | 0.01073 |
| MLFN AR(4)* | 0.00930 | 0.01152 |
| MLFN AR(5)* | 0.00984 | 0.01178 |
| MLFN AR(1)-MUIP(1)* | 0.00944 | 0.01149 |
| Transfer Fcn ARIMA(1,0,0)-MUIP(1/1)* | 0.00974 | 0.01199 |
| Transfer Fcn ARIMA(0,0,2)-MUIP(0/1)* | 0.01073 | 0.01332 |
| Transfer Fcn MUIP(0/1)* | 0.01046 | 0.01316 |
| Random Walk* | 0.01068 | 0.01375 |
| NN+Wavelet Denoising | 0.00481 | 0.00630 |
| NN+Wavelet Packet Denoising | 0.00485 | 0.00636 |
| New Method(DWT) | 0.00043 | 0.00054 |
| without statistical feature | 0.00407 | 0.00546 |
| New Method(SWT)† | 0.00377 | 0.00518 |

* The values in this row are taken from reference [69].
† The Mean Absolute Deviation and Mean features are used.

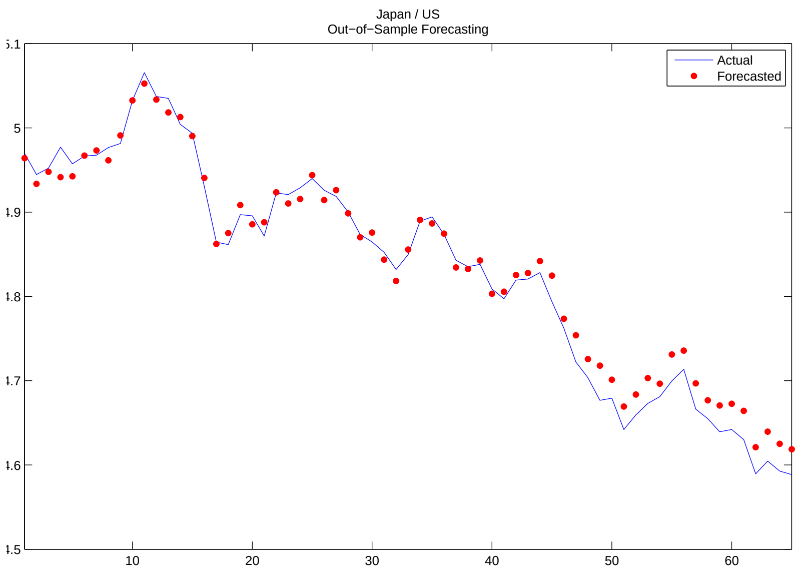



Figure 5.23: Japanese Yen/US Dollar

Table 5.25: Forecasting performance for Japanese Yen/US Dollar

| Method | MAE | RMSE |
|---|---|---|
| GRNN AR(1)* | 0.02119 | 0.02716 |
| GRNN AR(2)* | 0.02091 | 0.02687 |
| GRNN AR(12)* | 0.02152 | 0.02755 |
| MLFN AR(3)* | 0.02503 | 0.03159 |
| MLFN AR(4)* | 0.02480 | 0.03129 |
| MLFN AR(1)-MUIP(1)* | 0.02585 | 0.03264 |
| Transfer Fcn ARIMA(1,0,0)-MUIP(1/0)* | 0.02305 | 0.02925 |
| Transfer Fcn ARIMA(0,0,2)-MUIP(0/1)* | 0.02476 | 0.03071 |
| Transfer Fcn ARIMA(0,0,2)-MUIP(1/0)* | 0.02488 | 0.03093 |
| Random Walk* | 0.02667 | 0.03398 |
| NN+Wavelet Denoising | 0.02498 | 0.03150 |
| NN+Wavelet Packet Denoising | 0.05167 | 0.07260 |
| New Method(DWT) | 0.00397 | 0.00576 |
| without statistical feature | 0.01882 | 0.02295 |
| New Method(DWT)† | 0.01473 | 0.01847 |

* The values in this row are taken from reference [69].
† The Kurtosis and Mean features are used.

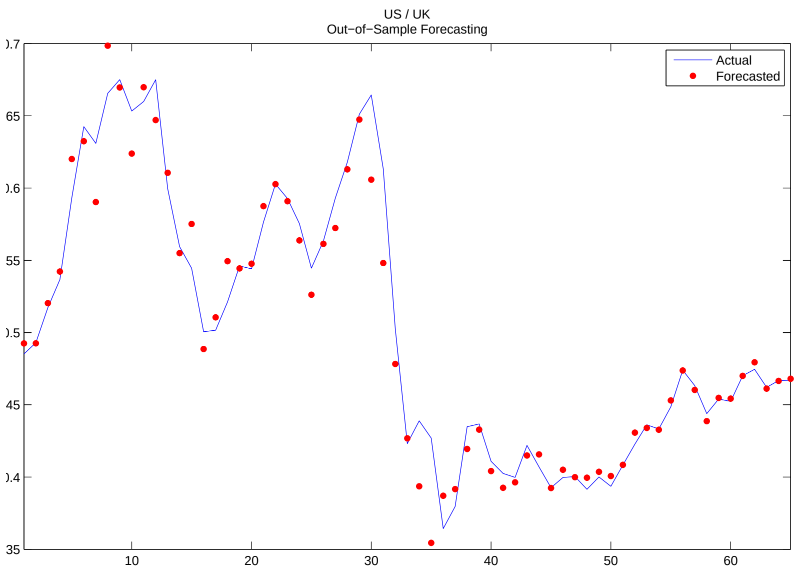



Figure 5.24: US Dollar/British Pound

Table 5.26: Forecasting performance for US Dollar/British Pound

| Method | MAE | RMSE |
|---|---|---|
| GRNN AR(1)* | 0.02212 | 0.02865 |
| GRNN AR(2)* | 0.02202 | 0.02855 |
| GRNN AR(3)-MUIP(1)* | 0.02161 | 0.02995 |
| MLFN AR(2)* | 0.02309 | 0.02996 |
| MLFN AR(3)* | 0.02468 | 0.03119 |
| MLFN AR(1)-MUIP(1)* | 0.02216 | 0.02990 |
| Transfer Fcn ARIMA(1,0,2)-MUIP(0/1)* | 0.02478 | 0.03200 |
| Transfer Fcn ARIMA(2,0,1)-MUIP(0/1)* | 0.02499 | 0.03174 |
| Transfer Fcn ARIMA(1,0,1)-MUIP(0/1)* | 0.02305 | 0.03141 |
| Random Walk* | 0.02706 | 0.03446 |
| NN+Wavelet Denoising | 0.01338 | 0.01655 |
| NN+Wavelet Packet Denoising | 0.01074 | 0.01388 |
| New Method(DWT) | 0.00169 | 0.00241 |
| without statistical feature | 0.01498 | 0.02753 |
| New Method(SWT)† | 0.01218 | 0.01989 |

* The values in this row are taken from reference [69].
† The Mean feature is used.

Table 5.27: Experiment settings five

| | |
|---|---|
| Exchange Rate(s) | Japan/US US/UK |
| Forecasting Horizon | One-Step-Ahead |
| Data Frequency | Monthly |
| Time Span | Jan-1971 to Dec-2003 |
| Training Set Size | 346 |
| Validation Set Size | 24 |
| Testing Set Size | 36 |
| Performance Measure(s) | NMSE |

Table 5.28: Forecasting performance for Japanese Yen/US Dollar and US Dollar/British Pound

| Method | NMSE | |
|---|---|---|
| | Japan/US | US/UK |
| GLAR* | 0.4207 | 0.0737 |
| ANN* | 0.1982 | 0.1031 |
| Hybrid* | 0.3854 | 0.0758 |
| EW* | 0.2573 | 0.0469 |
| ME* | 0.1505 | 0.0410 |
| NE* | 0.1391 | 0.0357 |
| NN+Wavelet Denoising | 0.1495 | 0.0767 |
| NN+Wavelet Packet Denoising | 0.0935 | 0.0933 |
| New Method(DWT) | 0.0006 | 0.0002 |
| without statistical feature | 0.0276 | 0.0274 |
| New Method(SWT) | 0.0262† | 0.0238‡ |

* The values in this row are taken from reference [62].
† The Mean Absolute Deviation and Skewness features are used.
‡ The Shannon Entropy and Turning Points features are used.

**Experiment Six**

In this experiment, we compared the performance of our system with the results from [8]. Below is a short description of [8] and the results are shown in Table 5.30 and Figure 5.27.

[8] compared the forecasting performance of a Takagi-Sugeno type neuro-fuzzy system and a feed-forward neural network that is trained using the scaled conjugate gradient algorithm. The results indicate that the proposed neuro-fuzzy method performed better than the neural network.

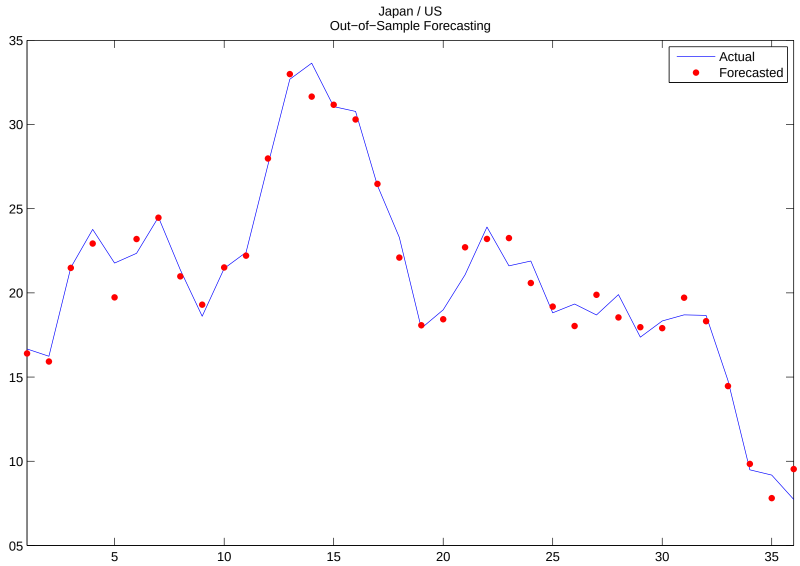



Figure 5.25: Japanese Yen/US Dollar

Table 5.29: Experiment settings six

| | |
|---|---|
| Exchange Rate(s) | US/Australia |
| Forecasting Horizon | One-Step-Ahead |
| Data Frequency | Monthly |
| Time Span | Jan-1981 to Apr-2001 |
| Training and Validation Sets Size | 170 |
| Testing Set Size | 73 |
| Performance Measure(s) | RMSE |

**Experiment Seven**

In this experiment, we compared the performance of our method with the approaches from [54]. Below is a short description of [54] and the results are shown in Table 5.32 and Figure 5.28.

[54] employs supervised neural networks as the metamodeling technique to design a financial time series forecasting system. The method works as follows: firstly, a cross-validation technique is used to generate different training subsets. Then, neural predictors with different initial conditions and training algorithms are trained to formulate forecasting base-models based on the different training subsets. Finally, a neural-network-based metamodel is obtained by learning from all base-models so as to improve the model accuracy. The performance of proposed approach is compared with several conventional forecasting models

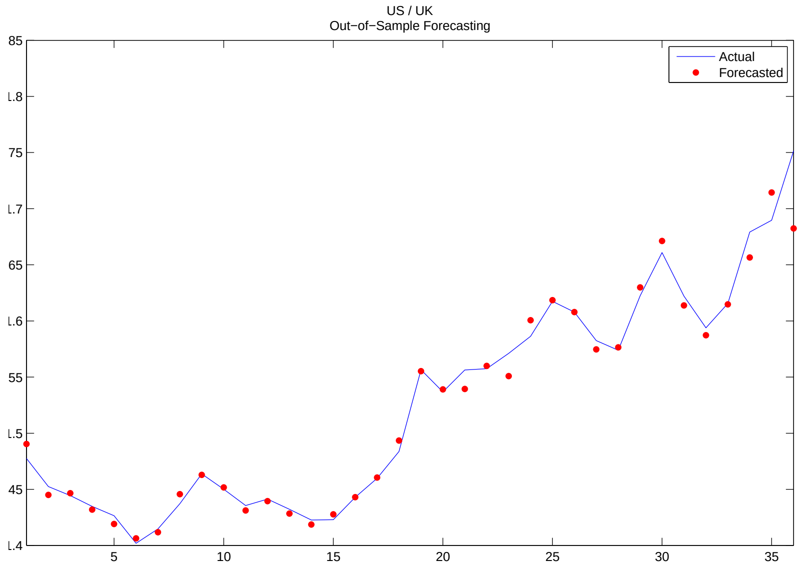



Figure 5.26: US Dollar/British Pound

Table 5.30: Forecasting performance for US Dollar/Australian Dollar

| Method | RMSE |
|---|---|
| Artificial neural network* | 0.034 |
| Neuro-Fuzzy system* | 0.034 |
| NN+Wavelet Denoising | 0.0104 |
| NN+Wavelet Packet Denoising | 0.0058 |
| New Method(DWT) | 0.0021 |
| without statistical feature | 0.0072 |
| New Method(SWT)† | 0.0052 |

* The values in this row are taken from reference [8].
† The Kurtosis and Mean features are used.

including Random Walk (RW) , Auto-Regression Integrated Moving Average (ARIMA), Exponential Smoothing (ES) and individual Back-Propagation Neural Network (BPNN) models. The results show that the neural-network-based metamodel outperforms all the other models that are tested in this study.

**Experiment Eight**

In this experiment, we compared the performance of our method with the approaches from [38]. Below is a short description of [38] and the results are shown in Table 5.34 and Figure 5.29.

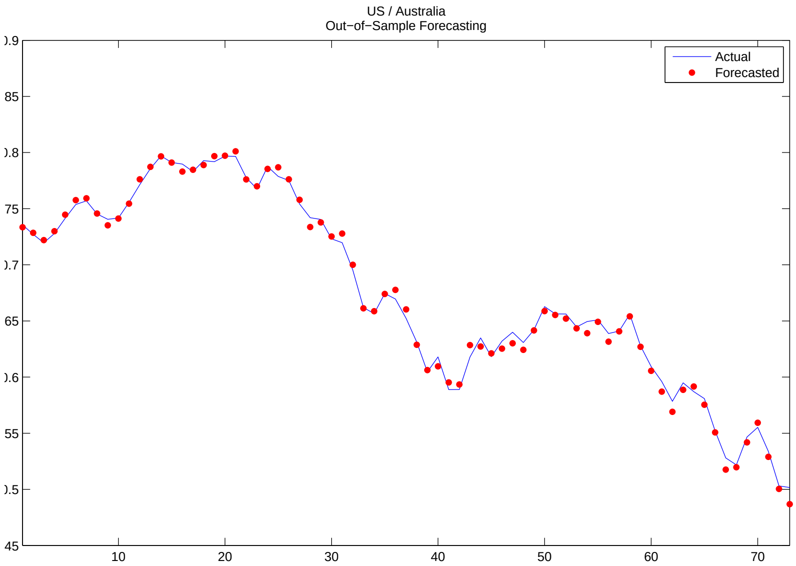



Figure 5.27: US Dollar/Australian Dollar

Table 5.31: Experiment settings seven

| | |
|---|---|
| Exchange Rate(s) | US/Euro |
| Forecasting Horizon | One-Step-Ahead |
| Data Frequency | Daily |
| Time Span | Jan-1-2000 to Dec-31-2004 |
| Training and Validation Sets Size | 1004 |
| Testing Set Size | 252 |
| Performance Measure(s) | RMSE |

[38] combines artificial neural networks (ANN) and economics to predict exchange rate fluctuations. More concretely,the paper includes a variable from the field of microeconomic, order flow (i.e. aggregated, small securities orders that brokers send to dealers often in return for cash payments), in a set of macroeconomic variables (interest rate and crude oil price) to help to explain the movements of exchange rate. Several forecasting models are tested including Linear/ANN Model 1 (Linear/ANN model + economic variables combination 1),Linear/ANN Model 2 (Linear/ANN model + economic variables combination 2), and Random Walk models. The empirical results indicate that ANN models outperform Random Walk and linear models in out-of-sample forecasts. And more importantly, the inclusion of macroeconomic and microeconomic variables improves the predictive power of both the linear and non-linear models.

Table 5.32: Forecasting performance for US Dollar/Euro

| Method | RMSE |
|---|---|
| RW* | 0.0066 |
| ARIMA* | 0.0041 |
| ES* | 0.0048 |
| BPNN* | 0.0052 |
| Metamodel* | 0.0038 |
| NN+Wavelet Denoising | 0.0065 |
| NN+Wavelet Packet Denoising | 0.0108 |
| New Method(DWT) | 0.0013 |
| without statistical feature | 0.0076 |
| New Method(SWT)† | 0.0026 |

* The values in this row are taken from reference [54].
† The Variance and Mean features are used.

Table 5.33: Experiment settings eight

| | |
|---|---|
| Exchange Rate(s) | Canada/US* |
| Forecasting Horizon | One-Step-Ahead, Seven-Step-Ahead |
| Data Frequency | Daily |
| Time Span | Jan-1990 to Jun-2000 |
| Training and Validation Sets Size | 2230 |
| Testing Set Size | 225 |
| Performance Measure(s) | RMSE |

* The natural logarithm of the exchange rates time series.

**Experiment Nine**

In this experiment, we compared the performance of our method with the results from [56]. Below is a short description of [56] and the results are shown in Tables 5.36,5.37 and Figures 5.30,5.31.

[56] proposed improved neural network and fuzzy models for exchange rate prediction that adopt the introduction of dynamics into the network by transferring the regular network neuron state to another set of duplication neurons called memory neurons. Furthermore, the one-step-ahead and multiple-step-ahead forecasting performance of several methods are tested, including Backpropagation (BP), Spread Encoding (SE), Radial Basis Functions (RBF), Autoregressive Recurrent Neural Network (ARNN),Elman Network (ELM), Neuro-Fuzzy system and Adaptive Fuzzy Logic System (AFLS). The experimental results demonstrate that the introduction of new algorithms can improve the accuracy of the prediction.

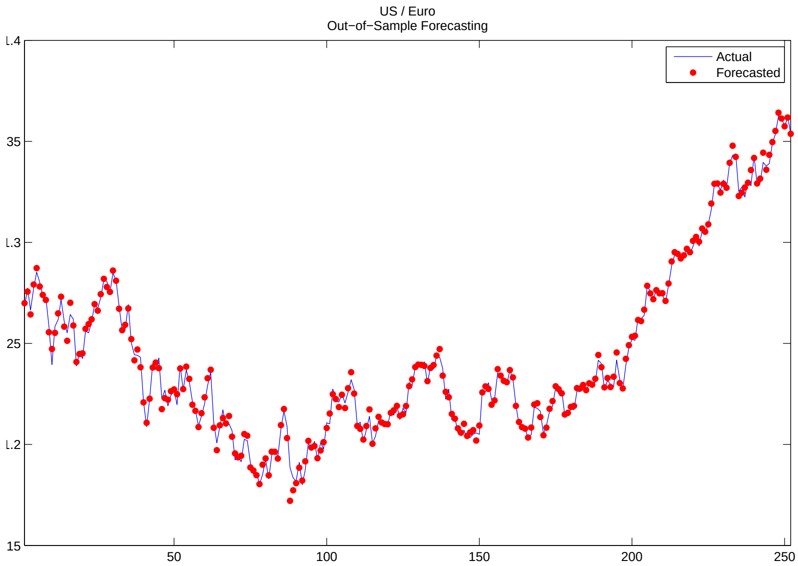



Figure 5.28: US Dollar/Euro

We can see from the results of the comparative experiments that the proposed hybrid model provides far more accurate forecasts than the approaches used in previous publications. Furthermore, the experimental results presented in Sections 5.2 and 5.3 indicate that firstly, the introduction of statistical features can considerably improve the out-of-sample forecasting performance of the model. Secondly, our method has apparent advantage over wavelet/wavelet packet denoising-based approaches in almost all the cases. And thirdly, the comparative performance between wavelet-denoising-based and wavelet-packet-denoising-based methods is quite mixed. Therefore, given an arbitrary foreign exchange rate time series, it is hard to say which approach will produce better forecasting results without doing experiments. As to the DWT version of the proposed model, although it gives amazingly good results especially in 1-day-ahead forecasting mode, we should not use it as the theoretical foundation of this method is not sound.

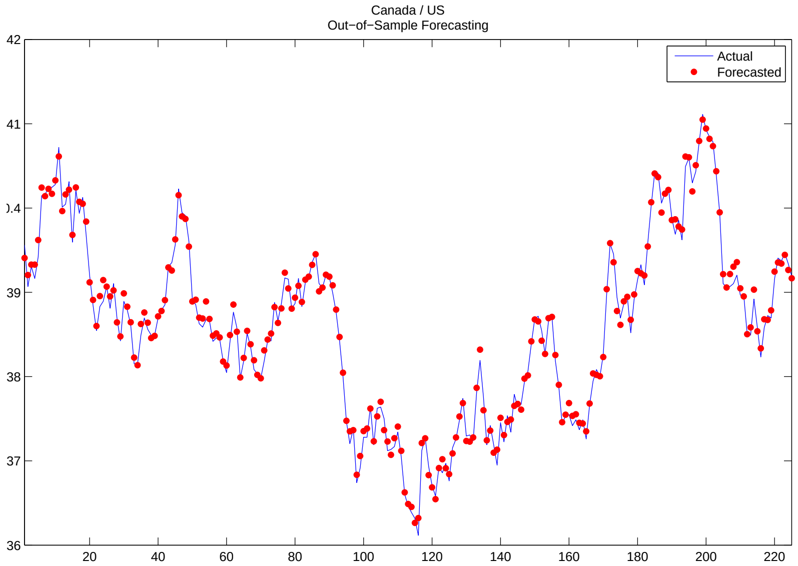
Table 5.34: Forecasting performance for Canadian Dollar/US Dollar

| Method | RMSE | |
|---|---|---|
| | 1-step-ahead | 7-step-ahead |
| Random Walk* | 0.0014321 | 0.003457 |
| Linear Model 1* | 0.0014364 | 0.003461 |
| ANN Model 1* | 0.0014270 | 0.003458 |
| Linear Model 2* | 0.0014331 | 0.003460 |
| ANN Model 2* | 0.0014216 | 0.003448 |
| NN+Wavelet Denoising | 0.0020590 | 0.017977 |
| NN+Wavelet Packet Denoising | 0.0021850 | 0.020346 |
| New Method(DWT) | 0.0005037 | 0.001811 |
| without statistical feature | 0.0009450 | 0.0028836 |
| New Method(SWT) | 0.0008258† | 0.0023099‡ |

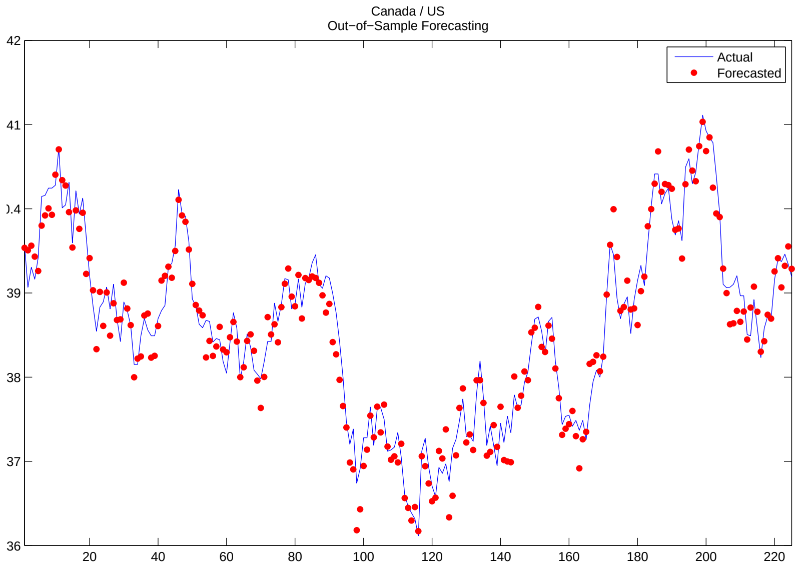
* The values in this row are taken from reference [38].
† The Mean Absolute Deviation and Mean features are used.
‡ The Kurtosis and Shannon Entropy features are used.

Table 5.35: Experiment settings nine

| | |
|---|---|
| Exchange Rate(s) | US/UK |
| Forecasting Horizon | One,Four,Eight,Twelve-Step-Ahead |
| Data Frequency | Daily |
| Time Span | Dec-31-1997 to 31-Mar-31-2000 |
| Training and Validation Sets Size | 800 |
| Testing Set Size | 200 |
| Performance Measure(s) | RMSE |




(a) 1-step-ahead




(b) 7-step-ahead

Figure 5.29: Forecasting Results for Canadian Dollar/US Dollar

Table 5.36: One-step-ahead Forecasting performance for US Dollar/British Pound

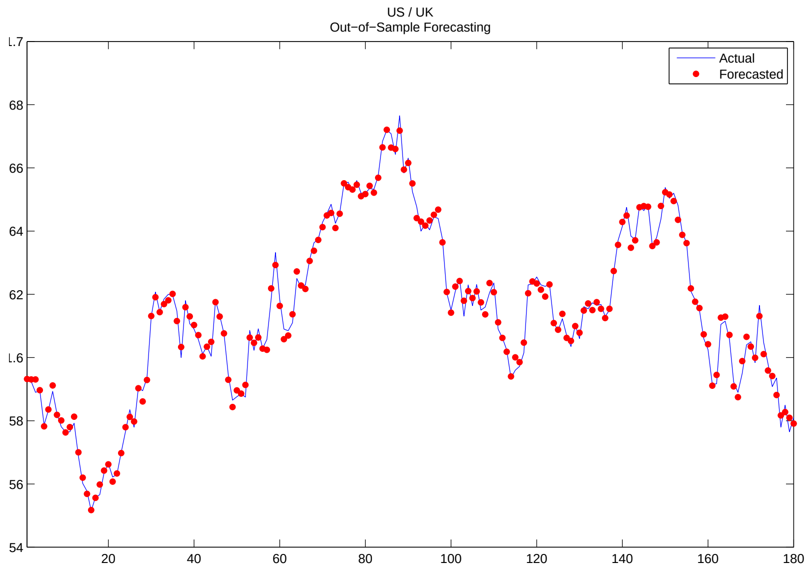
| Method | RMSE |
|---|---|
| BP* | 0.4215 |
| SE* | 0.4201 |
| RBF* | 0.3905 |
| ARNN* | 0.4021 |
| ELM* | 0.4010 |
| N-Fuzzy* | 0.39774 |
| AFLS/gain over BP* | 0.391 |
| NN+Wavelet Denoising | 0.00480 |
| NN+Wavelet Packet Denoising | 0.00858 |
| New Method(DWT) | 0.00017 |
| without statistical feature | 0.00242 |
| New Method(SWT)† | 0.00207 |

* The values in this row are taken from reference [56].
† The Mean Absolute Deviation and Shannon Entropy features are used.

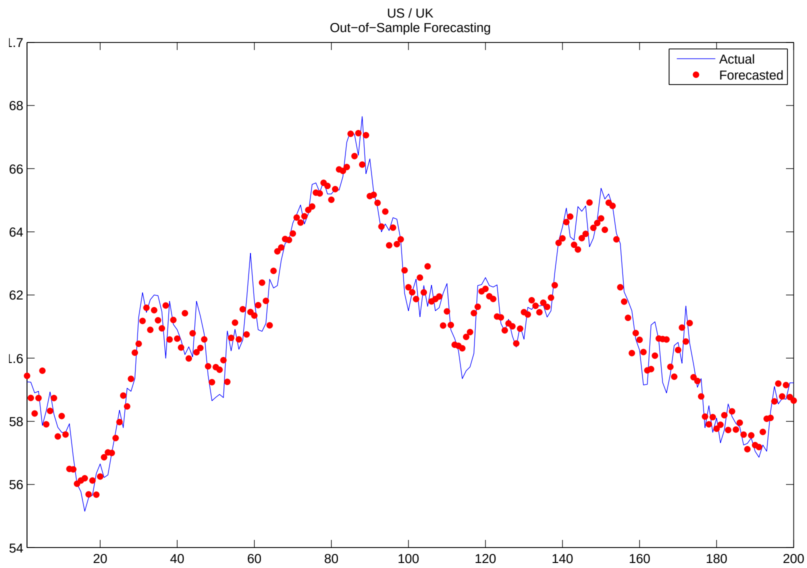
Table 5.37: Multiple-step-ahead Forecasting performance for US Dollar/British Pound

| Method | RMSE | | |
|---|---|---|---|
| | 4-step-ahead | 8-step-ahead | 12-step-ahead |
| BP* | 0.8952 | 1.1348 | 1.3120 |
| ELM* | 0.8505 | 1.0768 | 1.3196 |
| N-Fuzzy* | 0.8024 | 1.0498 | 1.2972 |
| AFLS/gain over BP* | 0.8067 | 1.0085 | 1.2597 |
| NN+Wavelet Denoising | 0.0064 | 0.0224 | 0.0320 |
| NN+Wavelet Packet Denoising | 0.0186 | 0.0335 | 0.0411 |
| New Method(DWT) | 0.0034 | 0.0048 | 0.0070 |
| without statistical feature | 0.0091 | 0.0106 | 0.0125 |
| New Method(SWT) | 0.0071† | 0.0078‡ | 0.0103× |

* The values in this row are taken from reference [56].
† The Shannon Entropy feature is used.
‡ The Variance and Mean features are used.
× The Shannon Entropy and Turning Points features are used.

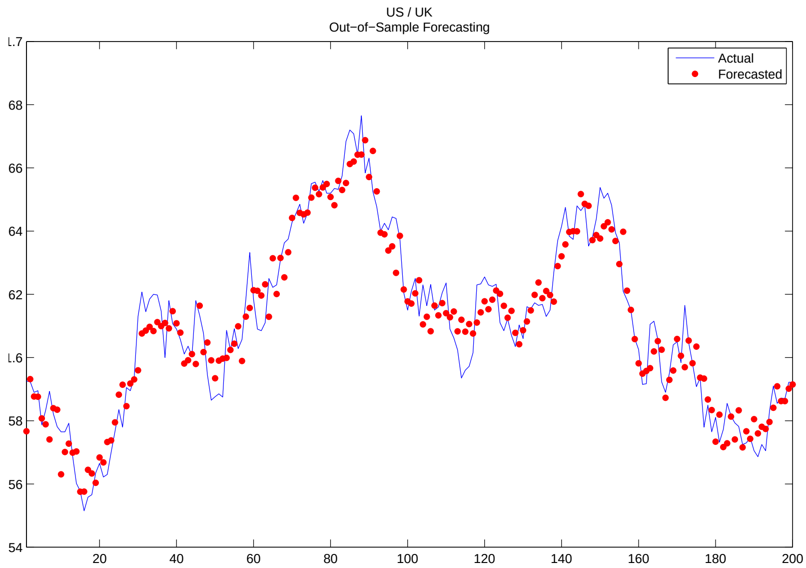



(a) 1-step-ahead

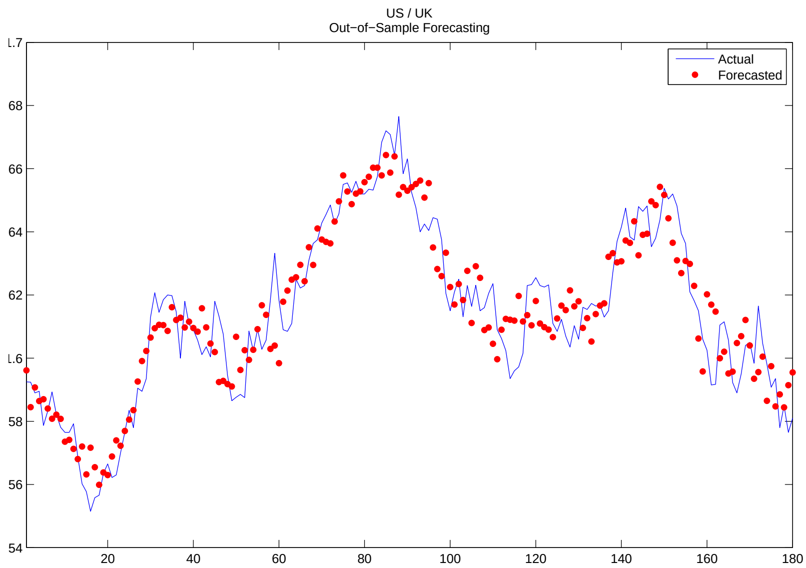



(b) 4-step-ahead

Figure 5.30: Forecasting Results for US Dollar/British Pound (1)




(a) 8-step-ahead




(b) 12-step-ahead

Figure 5.31: Forecasting Results for US Dollar/British Pound (2)

# Chapter 6

## Conclusion and Future Work

In this thesis, we presented an improved forecasting model based on neural network, stationary wavelet transform and statistical time series analysis techniques. We compared the new model's performance with pure neural network forecasting model, wavelet/wavelet-packet-denoising-based forecasting models and the approaches used in nine previous studies. Moreover, we illustrated the impact of using non-shift-invariant discrete wavelet transform on the accuracy of forecasting. Our experimental results indicate that the proposed model provides much better forecasting results than existing methods. Therefore, the work demonstrates the feasibility of combining wavelet neural network and statistical methods to achieve accurate forecasting.

Our future work includes the following tasks: 1. using more economically significant features, such as oil prices, import and export values, interest rates and the growth rates of GDP(gross domestic product), to enhance the accuracy of forecasting. 2. trying to improve the performance of the neural networks in the system using fuzzy logic and evolutionary algorithms. 3. applying the proposed model to share price indices, interest rates and other financial time series. 4. developing a more sophisticated feature selection strategy to optimize the efficiency of the system. 5. improving the model's middle-term and long-term forecasting performance.

# Appendix A

## Appendix

### A.1 Statistical Performance Measures

In this study, we employ several measures to evaluate the performance of different forecasting models.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{A.1}$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} \tag{A.2}$$

Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \tag{A.3}$$

Root Mean Squared Error (RMSE)

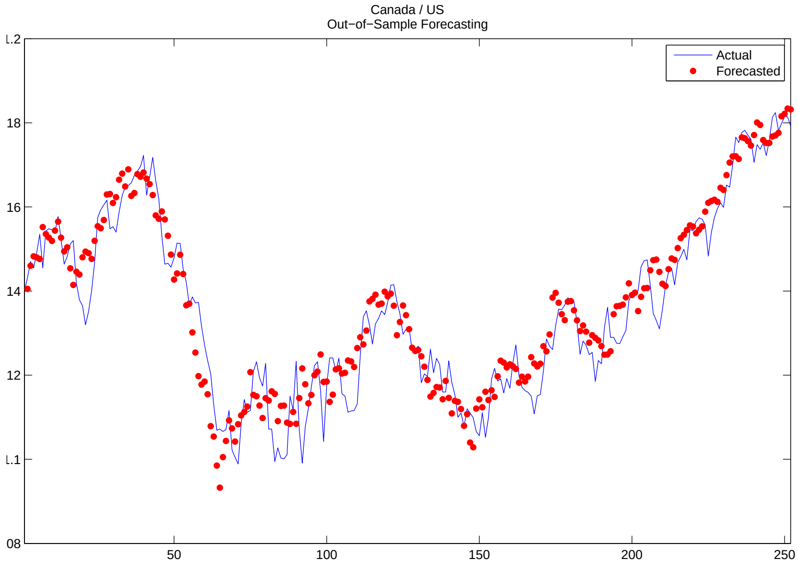
$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}} \tag{A.4}$$

Normalized Mean Square Error (NMSE)

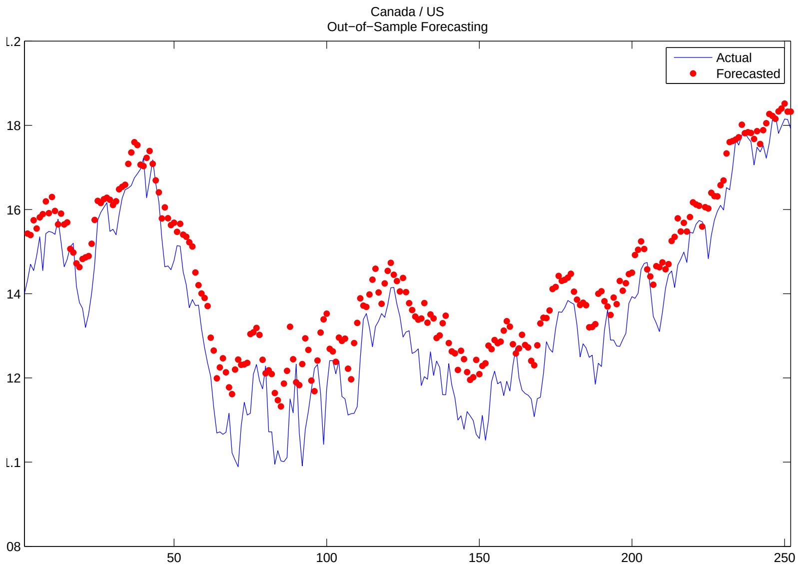
$$NMSE = \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y}_t)^2} \tag{A.5}$$

where $y_t$ and $\hat{y}_t$ are the actual and predicted values, $\bar{y}_t$ is the mean value of $y_t$, $t = 1...n$. The smaller values of error, the closer are the predicted values to the actual values.

## A.2 Middle-term and Long-term Forecasting Results

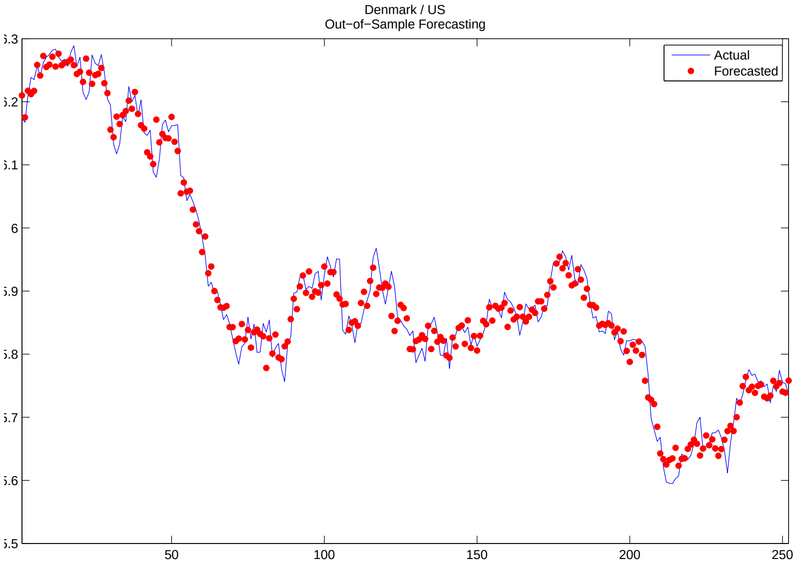



(a) 5-step-ahead

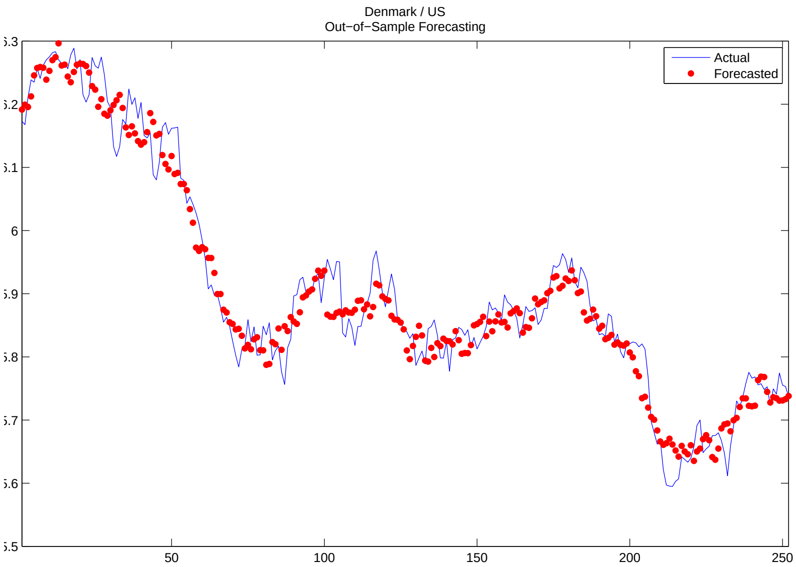



(b) 10-step-ahead

Figure A.1: Middle-term and Long-term Forecasting Results for Canadian Dollar/US Dollar

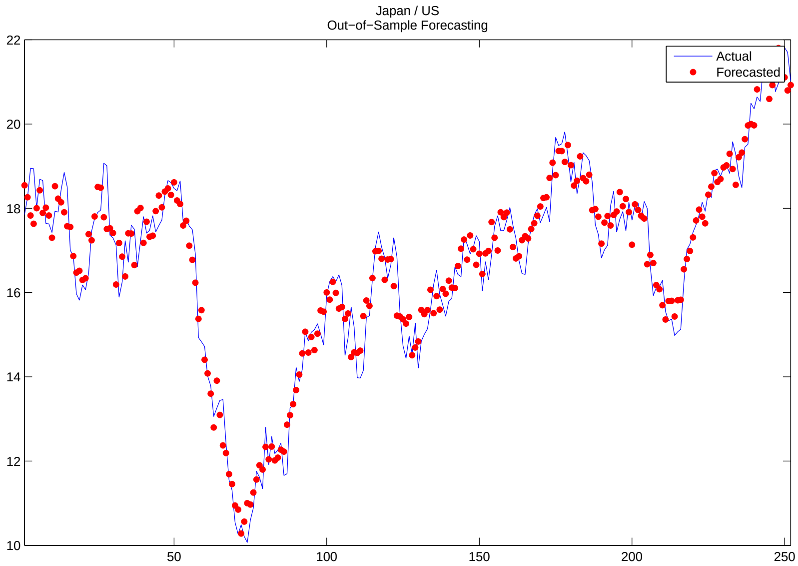



(a) 5-step-ahead

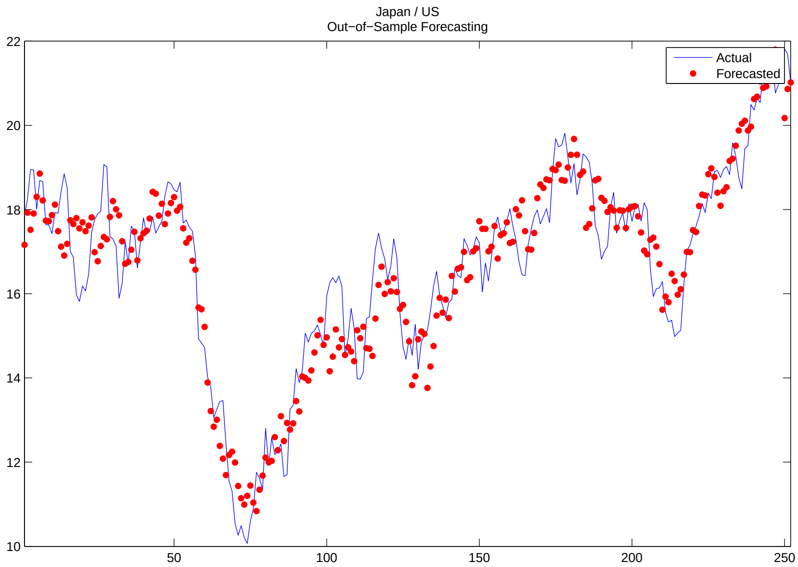



(b) 10-step-ahead

Figure A.2: Middle-term and Long-term Forecasting Results for Danish Kroner/US Dollar

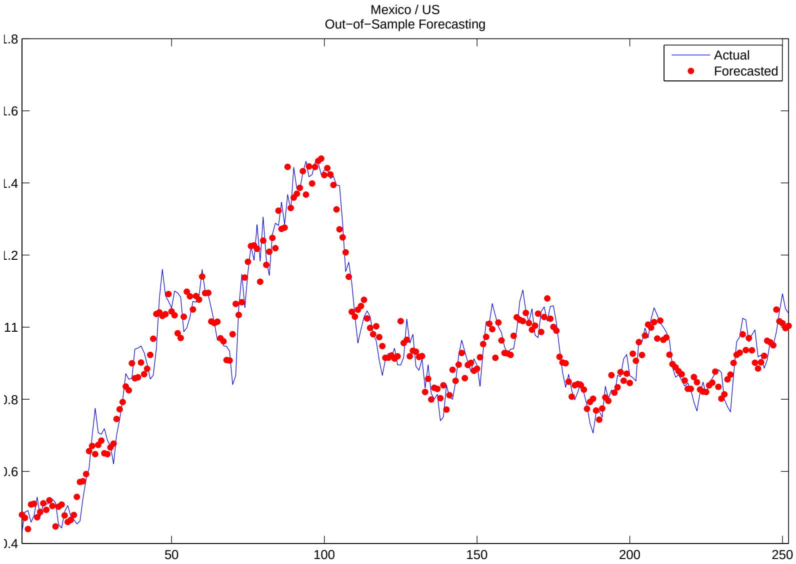



(a) 5-step-ahead

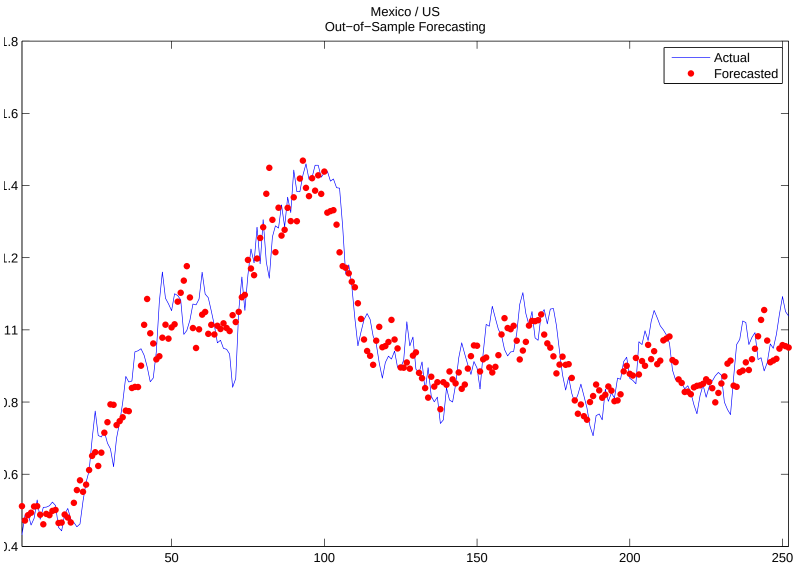



(b) 10-step-ahead

Figure A.3: Middle-term and Long-term Forecasting Results for Japanese Yen/US Dollar

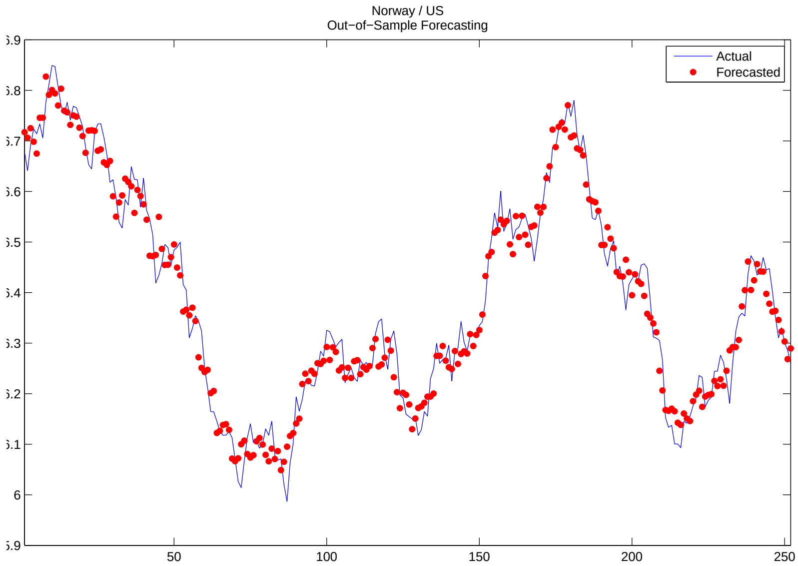



(a) 5-step-ahead

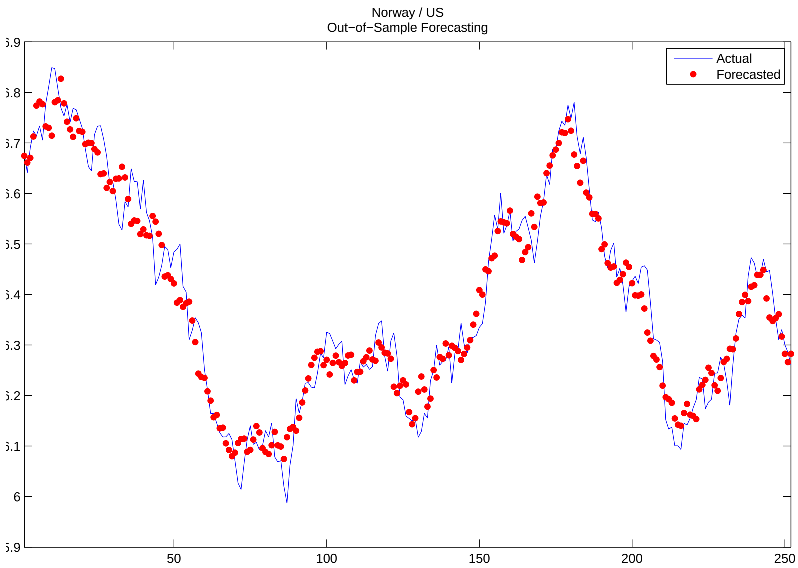



(b) 10-step-ahead

Figure A.4: Middle-term and Long-term Forecasting Results for Mexican New Pesos/US Dollar

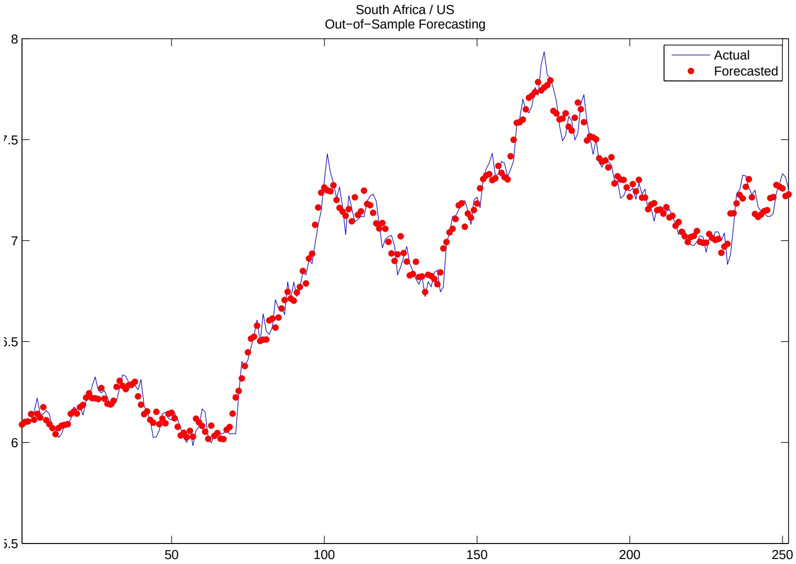



(a) 5-step-ahead

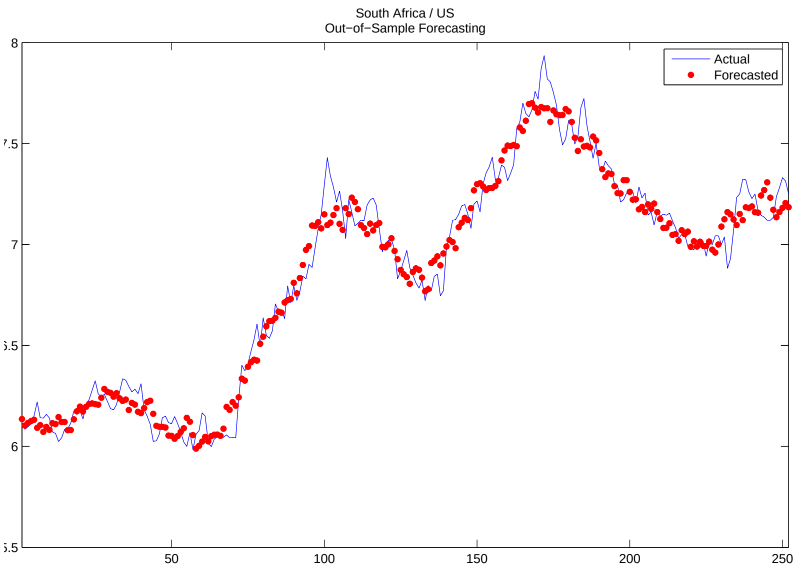



(b) 10-step-ahead

Figure A.5: Middle-term and Long-term Forecasting Results for Norwegian Kroner/US Dollar

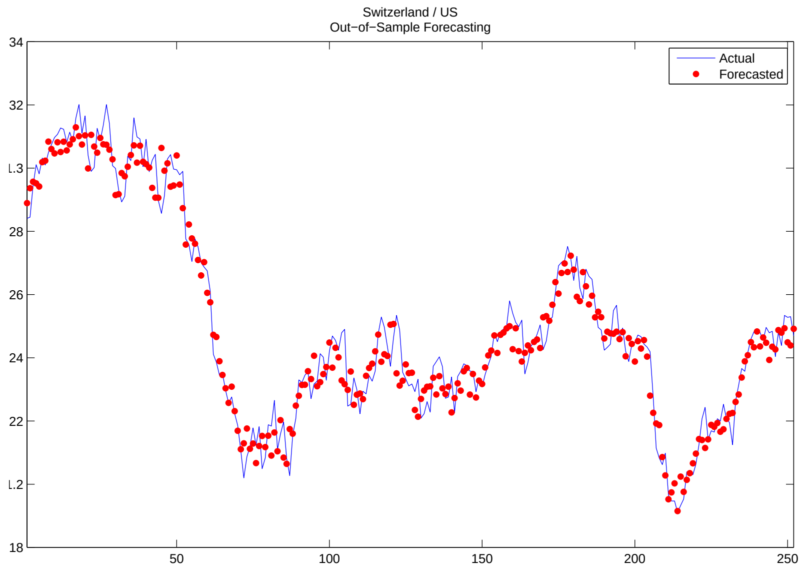



(a) 5-step-ahead

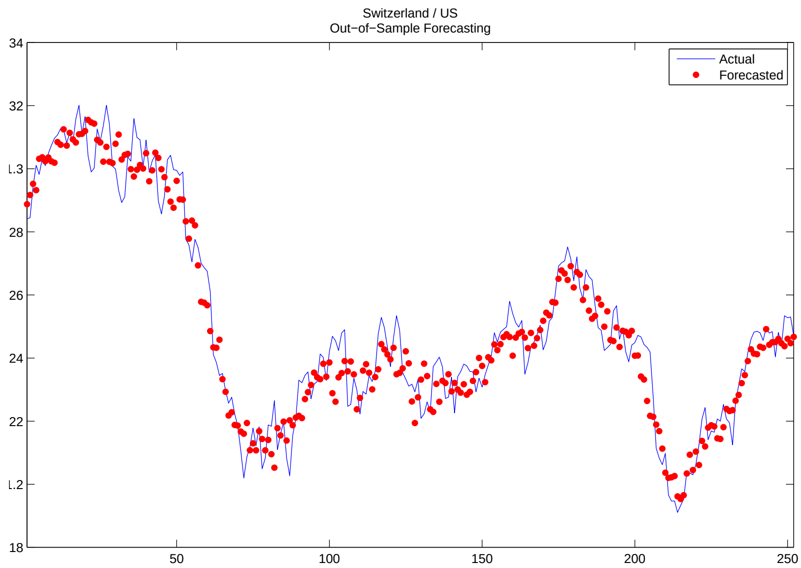



(b) 10-step-ahead

Figure A.6: Middle-term and Long-term Forecasting Results for South African Rand/US Dollar

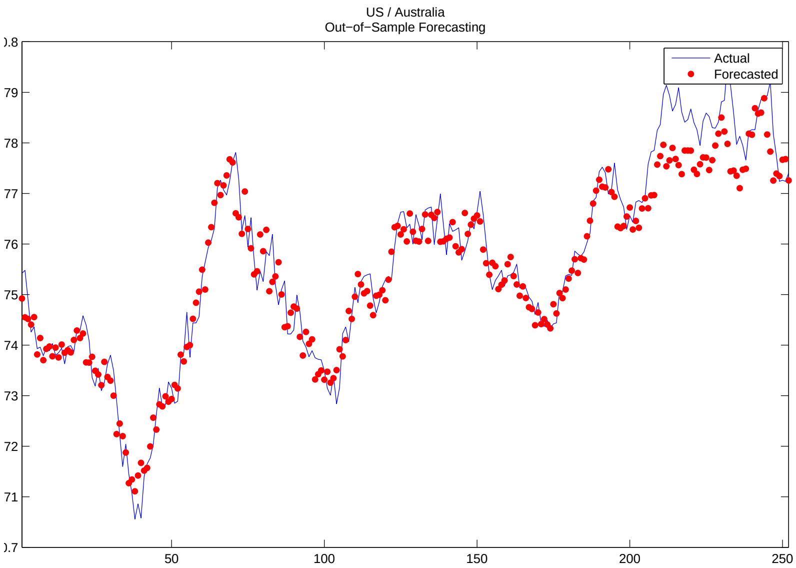



(a) 5-step-ahead

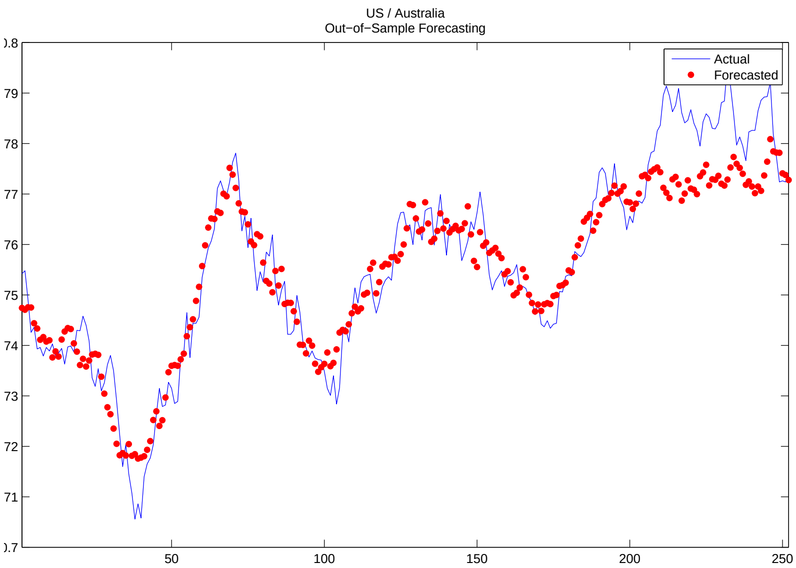



(b) 10-step-ahead

Figure A.7: Middle-term and Long-term Forecasting Results for Swiss Francs/US Dollar

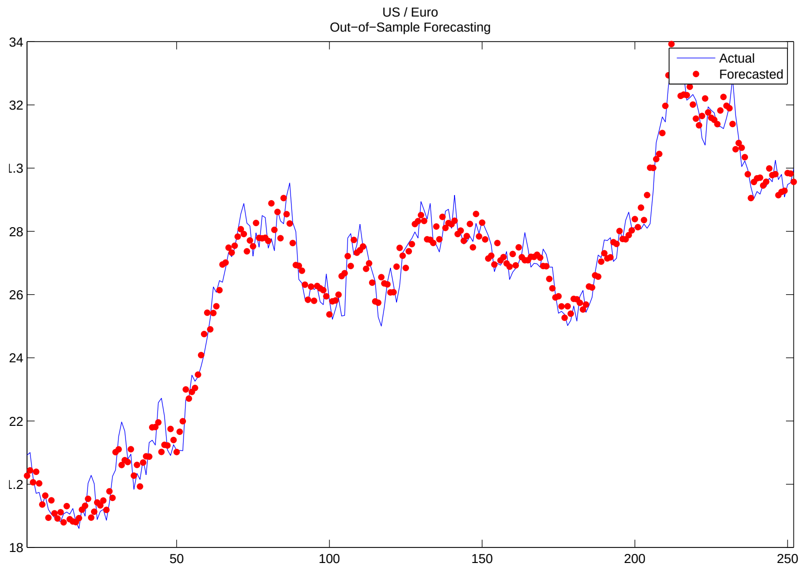



(a) 5-step-ahead

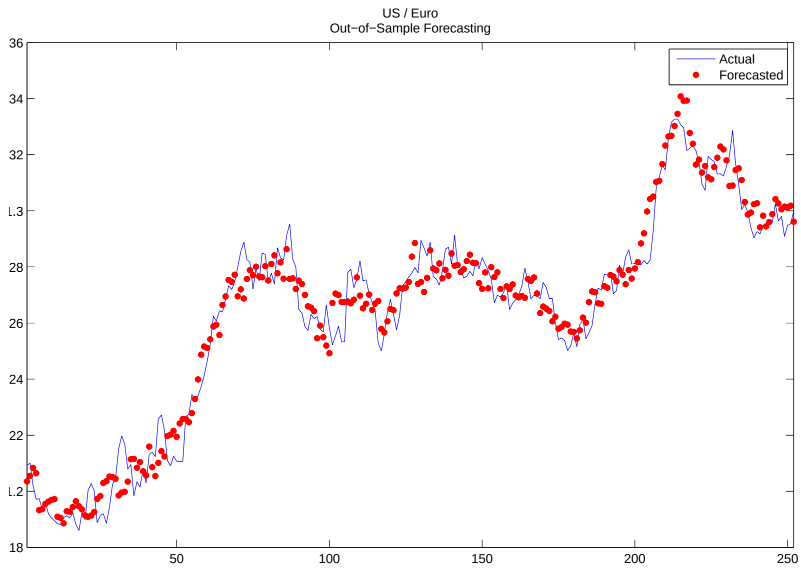



(b) 10-step-ahead

Figure A.8: Middle-term and Long-term Forecasting Results for US Dollar/Australian Dollar

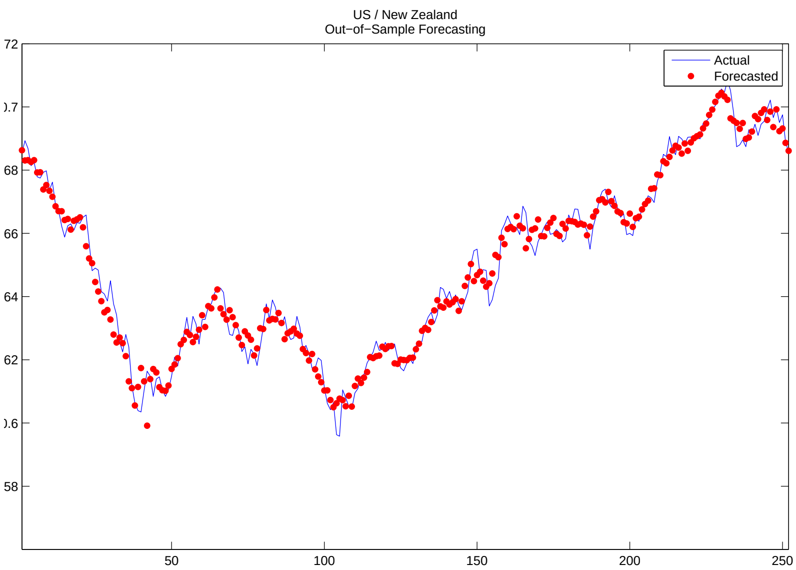



(a) 5-step-ahead

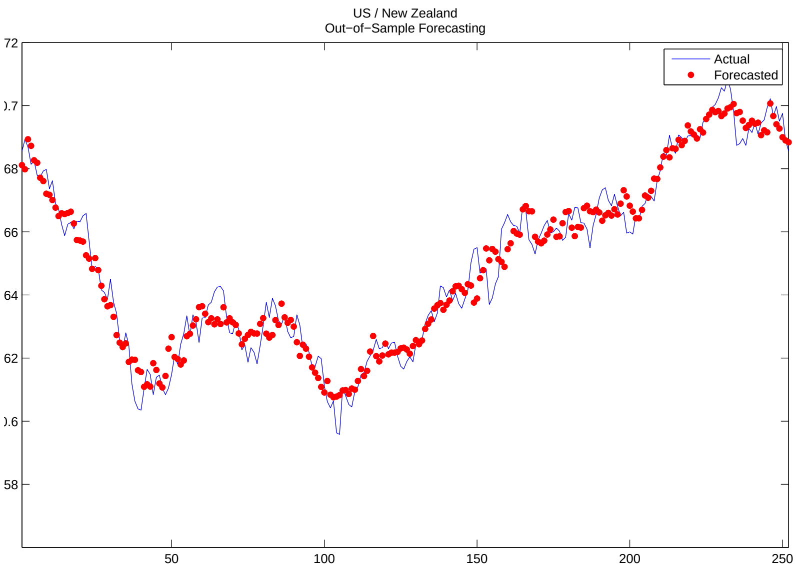



(b) 10-step-ahead

Figure A.9: Middle-term and Long-term Forecasting Results for US Dollar/Euro

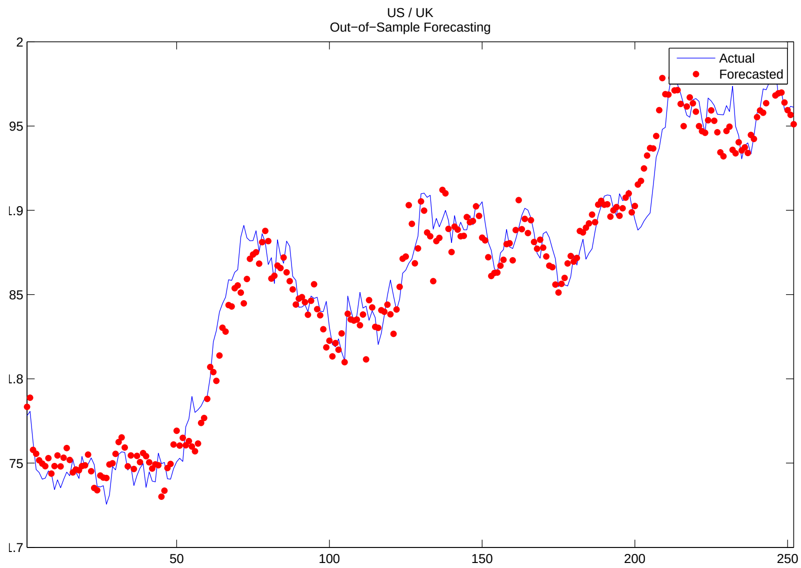



(a) 5-step-ahead

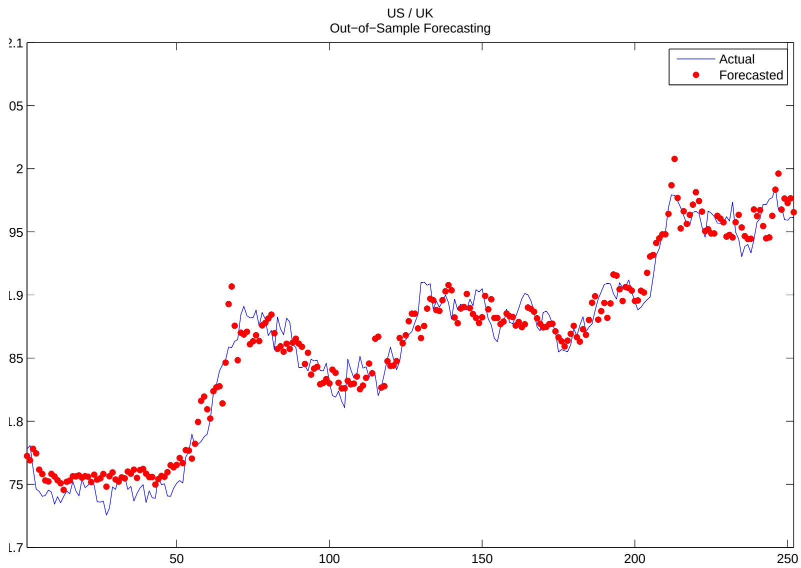



(b) 10-step-ahead

Figure A.10: Middle-term and Long-term Forecasting Results for US Dollar/New Zealand Dollar




(a) 5-step-ahead




(b) 10-step-ahead

Figure A.11: Middle-term and Long-term Forecasting Results for US Dollar/British Pound

# Bibliography


[1] exchange rate. Available at http://en.wikipedia.org/wiki/Exchange_rate.

[2] Foreign exchange. Available at http://www.exinfm.com/training/63011l7.rtf.

[3] Matlab - the language of technical computing. Available at http://www.mathworks.com/products/matlab/.

[4] The nan-toolbox: A statistic-toolbox for octave and matlab. Available at http://hci.tugraz.at/schloegl/matlab/NaN/.

[5] Netlab neural network software. Available at http://www.ncrg.aston.ac.uk/netlab/index.php.

[6] Remarks by chairman alan greenspan. Available at http://www.federalreserve.gov/BOARDDOCS/SPEECHES/2004/20041119/default.htm#fn2.

[7] St. louis fed: Exchange rates. Available at http://research.stlouisfed.org/fred2/categories/15.

[8] M. U. C. A Abraham and S. Petrovic-lazarevic. Australian forex market analysis using connectionist models. *Management Journal of Management Theory and Practice*, 29:18–22, 2003.

[9] J. C. A Aussem and F. Murtagh. Wavelet-based feature extraction and decomposition strategies for financial forecasting. *Journal of Computational Intelligence in Finance*, 6:5–12, 1998.

[10] P. G. Allen and B. J. Morzuch. Twenty-five years of progress, problems, and conflicting evidence in econometric forecasting. what about the next 25 years? *International Journal of Forecasting*, 22:475–492, 2006.

[11] L. Anastasakis and N. Mort. Applying a feedforward neural network for the prediction of the usd/gbp exchange rate. *In Proceedings of the 3rd Conference on Technology and Automation*, 1:169–174, 2000.

[12] D. P. A. Ashwani Kumar and S. D. Joshi. Study of canada/us dollar exchange rate movements using recurrent neural network model of fx-market. *IDA 2003, Lecture Notes in Computer Science*, 2779:409–417.

[13] G. Benavides and C. Capistran. Forecasting exchange rate volatility: The superior performance of conditional combinations of time series and option implied forecasts. *Banco de Mexico, working paper*, 1, 2009.

[14] U. L.-W. Bernd Brandl and S. Pickl. Increasing the fitness of fundamental exchange rate forecast models. *Int. J. Contemp. Math. Sciences*, 4(16):779–798, 2009.

[15] A. P. Bradley. Shift-invariance in the discrete wavelet transform. *Proceedings of the Seventh International Conference on Digital Image Computing: Techniques and Applications*, 1:29–38, 2003.

[16] S. L. C Lee Giles and A. C. Tsoi. Noisy time series prediction using a recurrent neural network and grammatical inference. *Machine Learning*, pages 161–183, 2001.

[17] S. G. Catherine Vairappan, Hiroki Tamura and Z. Tang. Batch type local search-based adaptive neuro-fuzzy inference system (anfis) with self-feedbacks for time-series prediction. *Neurocomputing*, 72:1870–1877, 2009.

[18] J. R. Chan Wing H and K. Madhu. The economic value of using realized volatility in forecasting future implied volatility. *Journal of Financial Research*, forthcoming, 2009.

[19] C.-T. Chang. A modified goal programming approach for the mean-absolute deviation portfolio optimization model. *Applied Mathematics and Computation*, 171:567–572, 2005.

[20] A.-S. Chena and M. T. Leung. Regression neural network for error correction in foreign exchange forecasting and trading. *Computers & Operations Research*, 31:1049–1068, 2004.

[21] K. B. Cho and B. H. Wang. Radial basis function based adaptive fuzzy systems their application to system identification and prediction. *Fuzzy Sets and Systems*, 83:325–339, 1995.

[22] C. K. Chui. *Wavelets: A Mathematical Tool for Signal Analysis*. Society for Industrial Mathematics, 1987.

[23] A. C. S. J. E. J. U. P. C. N. D B de Araujo, W Tedeschi and O. Baffa. Shannon entropy applied to the analysis of event-related fmri time series. *NeuroImage*, 20:311–317, 2003.

[24] G. E. H. D. E. Rumelhart and R. J. Williams. Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1*, pages 318–362, 1986.

[25] C. K. D Kim. Forecasting time series with genetic fuzzy predictor ensembles. *IEEE Transactions on Fuzzy Systems*, 5:523–535, 1997.

[26] I. Daubechies. *Ten Lectures Of Wavelets*. SIAM Press, 1992.

[27] P. Domingos. The role of occam's razor in knowledge discovery. *Data Mining and Knowledge Discovery*, 3:409–425, 1999.

[28] D. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224, 1995.

[29] D. L. Donoho. Denoising via soft thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.

[30] H. Du and N. Zhang. Time series prediction using evolving radial basis function networks with new encoding scheme. *Neurocomputing*, 71:1388–1400, 2008.

[31] C. L. Dunis and M. Williams. Modelling and trading the eur/usd exchange rate: Do neural network models perform better? *Derivatives Use, Trading & Regulation*, 8:211–239, 2002.

[32] A. B. Ercan Balaban and R. W. Faff. Forecasting stock market volatility: Further international evidence. *European Journal of Finance*, 12:171–188, 2006.

[33] D. B. Eva Andersson and M. Frisén. Detection of turning points in business cycles. *Journal of Business Cycle Measurement and Analysis*, 1, 2004.

[34] D. Fay and J. V. Ringwood. A wavelet transfer model for time series forecasting. *International Journal of Bifurcation and Chaos*, 17:3691–3696, 2007.

[35] P. N. Franck Moraux and C. Villa. The predictive power of the french market volatility index: A multi horizons study. *European Finance Review*, 2:303–320, 1999.

[36] D. Gnanadurai and V. Sadasivam. An efficient adaptive thresholding technique for wavelet based image denoising. *International Journal of Signal Processing*, 2:114–120, 2006.

[37] J. G. D. Gooijer and R. J. Hyndman. 25 years of time series forecasting. *International Journal of Forecasting*, 22:443–473, 2006.

[38] N. Gradojevic and J. Yang. The application of artificial neural networks to exchange rate forecasting: The role of market microstructure variables. *Bank of Canada Working Paper*, 23, 2000.

[39] M. Grigoletto and F. Lisi. Looking for skewness in financial time series. *XLIII Riunione Scientifica della Societa Italiana di Statistica*, 2006.

[40] S. Guoxiang and Z. Ruizhen. Three novel models of threshold estimator for wavelet coefficients. *WAA 2001, Lecture Notes In Computer Science*, 2251:145–150, 2001.

[41] D. Harding. Using turning point information to study economic dynamics. *Econometric Society 2004 Australasian Meetings, Econometric Society*, 2004.

[42] J. L. B. J. O. B. P. F. J. P. I Rojas, H Pomares and A. Prieto. Time series analysis using normalized pg-rbf network with regression weights. *Neurocomputing*, 42:267–285, 2002.

[43] H. P. L. J. H. J. G. Ignacio Rojas, Fernando Rojas and O. Valenzuela. The synergy between classical and soft-computing techniques for time series prediction. *Lecture notes in computer science vol. 2972*, pages 30–39, 2004.

[44] L. F. Jacques Anas, Monica Billio, , and G. L. Mazzi. A system for dating and detecting turning points in the euro area. *Manchester School*, 76:549–577, 2008.

[45] J.-S. R. Jang and C.-T. Sun. Predicting chaotic time series with fuzzy if-then rules. *Second IEEE International Conference on Fuzzy Systems*, 2:1079–1084, 1993.

[46] D. D. Jensen and P. R. Cohen. Multiple comparisons in induction algorithms. *Machine Learning*, 38(3):309–338, 2000.

[47] J. H. John Beirne and M. Simpson. Is the real exchange rate stationary? the application of similar tests for a unit root in the univariate and panel cases. *Quantitative and Qualitative Analysis in Social Science*, 1:55–70, 2007.

[48] J.P.Morgan/Reuters. Riskmetrics - technical document, fourth edition. *Morgan Guaranty Trust Company of New York, USA*, 1996.

[49] L. K and W. J. G. Predicting cyclical turning points with a leading index in a markov switching model. *Journal of Forecasting*, 13:245–263, 1994.

[50] J. F. Kaashoek and H. K. van Dijk. Neural network pruning applied to real exchange rate analysis. *Journal of Forecasting*, 21:559–577, 2002.

[51] J. Kamruzzaman and R. A. Sarker. Forecasting of currency exchange rates using ann: a case study. *Proceedings of the 2003 International Conference on Neural Networks and Signal Processing*, 1:793–797, 2003.

[52] F. Keinert. Wavelets and multiwavelets, studies in advanced mathematics. *Chapman & Hall/CRC Press.*

[53] I. Kim and S.-H. Lee. Open loop prediction of a time series by using fuzzy rules. *Proeedings of KFMS Spring Conference*, pages 290–295, 1994.

[54] S. W. Kin Keung Lai, Lean Yu and C. Zhou. Neural-network-based metamodeling for financial time series forecasting. *Proceedings of the 2006 Joint Conference on Information Sciences, JCIS 2006*, 2006.

[55] W. H. Kin Keung Lai, Lean Yu and S. Wang. Multistage neural network metalearning with application to foreign exchange rates forecasting. *Lecture Notes in Computer Science*, 4293:338–347, 2006.

[56] V. Kodogiannis and A. Lolis. Forecasting financial time series using neural network and fuzzy system-based techniques. *Neural Computing & Applications*, 11:90–102, 2002.

[57] L. W. Kok Keong Teo and Z. Lin. Wavelet packet multi-layer perceptron for chaotic time series prediction: Effects of weight initialization. *Lecture Notes in Computer Science*, 2074:310–317, 2001.

[58] H. Konno and T. Koshizuka. Mean-absolute deviation model. *IIE Transactions*, 37:893–900, 2005.

[59] A. M. Kristiaan Kerstens and I. V. de Woestyne. Geometric representation of the mean-variance-skewness portfolio frontier based upon the shortage function. *Tenth European Workshop on Efficiency and Productivity Measurement*, 2007.

[60] D. C. L and H. Xuehuan. Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *Journal of Forecasting*, 21:317–354, 2002.

[61] K. Lahiri and G. H. Moore. *Leading economic indicators: New approaches and forecasting records.* Cambridge University Press, Cambridge, UK, 1992.

[62] S. W. Lean Yu and K. K. Lai. A novel nonlinear ensemble forecasting model incorporating glar andann for foreign exchange rates. *Computers & Operations Research*, 32:2523–2541, 2005.

[63] W. H. Lean Yu, Shouyang Wang and K. K. Lai. Are foreign exchange rates predictable? a survey from artificial neural networks perspective. *Scientific Inquiry*, 8(2):207–228, 2007.

[64] S.-H. Lee and I. Kim. Time series analysis using fuzzy learning. *Proceedings of International Conference on Neural Information Processing*, 6:1577–1582, 1994.

[65] J. M. LI-XIN WANG. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems Man and Cybernetics*, 22:1414–1427, 1992.

[66] F. Lisi and R. A. Schiavo. A comparison between neural networks and chaotic models for exchange rate prediction. *Computational Statistics & Data Analysis*, 30:87–102, 1999.

[67] T. Lux and T. Kaizoji. Forecasting volatility and volume in the tokyo stock market: The advantage of long memory models. *Economics Working Papers*, 5, 2004.

[68] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[69] A.-S. C. Mark T Leung and H. Daouk. Forecasting exchange rates using general regression neural networks. *Computers & Operations Research*, 27:1093–1110, 2000.

[70] R. Marschinski and L. Matassini. Financial markets as a complex system: A short time scale perspective. *Deutsche Bank Research Notes in Economics& Statistics*, RN-01-4, November, 2001.

[71] R. Martin and B. Heinrich. A direct adaptive method for faster backpropagation learning: the rprop algorithm. *Proceedings of the IEEE International Conference on Neural Networks 1993*, pages 586–591.

[72] Minsky and Papert. Perceptrons - an introduction to computational geometry. *The MIT press, London, England*, 1969.

[73] U. Z. Miron Kaufman and P. S. Sung. Entropy of electromyography time series. *Physica A: Statistical Mechanics and its Applications*, 386:698–707, 2007.

[74] L. Molgedey and W. Ebeling. Local order, entropy and predictability of financial time series. *European Physical Journal B: Condensed Matter and Complex Systems*, 15:733–737, 2000.

[75] M. F. Møller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6:525–533, 1993.

[76] A. K. Nag and A. Mitra. Forecasting daily foreign exchange rates using genetically optimized neural networks. *Journal of Forecasting*, 21:501–511, 2002.

[77] S. Natenberg. Option volatility and pricing strategies: Advanced trading techniques for professionals. Probus Publishing Company, 1988.

[78] L. S. O Renaud and F. Murtagh. Wavelet-based forecasting of short and long memory time series. *Cahiers du Departement d'Econometrie*, 4, 2002.

[79] T. K. R. P Jana and S. K. Majumdar. Multi-objective mean-variance-skewness model for portfolio optimization. *AMO - Advanced Modeling and Optimization*, 9:181–193, 2007.

[80] W.-C. H. Ping-Feng Pai, Chih-Shen Lin and C.-T. Chen. A hybrid support vector machine regression for exchange rate prediction. *Information and Management Sciences*, 17(2):19–32, 2006.

[81] E. Plummer. Time series forecasting with feed-forward neural networks: Guidelines and limitations. Available at `http://www.karlbranting.net/papers/plummer/Pres.ppt`.

[82] S.-H. Poon and C. Granger. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, XLI:478–539, 2003.

[83] M. Qi and Y. Wu. Nonlinear prediction of exchange rates with monetary fundamentals. *Journal of Empirical Finance*, 10:623–640, 2003.

[84] J. B. Ramsey. Wavelets in economics and finance: Past and future. *Working Paper No.S-MF-02-02, C.V. Starr Center for Applied Economics, New York University*, 2002.

[85] S. Z. M. H. Roselina Sallehuddin, Siti Mariyam Hj Shamsuddin and A. Abraham. Forecasting time series data using hybrid grey relational artificial nn and auto-regressive integrated moving average model. *Neural Network World*, 6:573–605, 2007.

[86] M. J. L. Rumelhart D E and T. P. R. Group. Parallel distributed processing: Explorations in the microstructure of cognition. *The MIT Press,Cambridge, Mass*, 1986.

[87] R. K. P. S Velickov and D. P. Solomatine. Prediction of nonlinear dynamical systems based on time series analysis: Issues of entropy, complexity and predictability. *Proceedings of the XXX IAHR Congress, Thessaloniki, Greece*, August,2003.

[88] N. Scafetta and B. J. West. Multiresolution diffusion entropy analysis of time series: an application to births to teenagers in texas. *Chaos, Solitons and Fractals*, 20:179–185, 2004.

[89] G. Serna. On the information content of volatility, skewness and kurtosis implied in option prices. *Documentos de Trabajo, Universidad de Castilla La Mancha*, 1, 2001.

[90] I. W. Shahriar Yousefi and D. Reinarz. Wavelet-based prediction of oil prices. *Chaos, Solitons and Fractals*, 25:265–275, 2005.

[91] S. Singh and J. Fieldsend. Pattern matching and neural networks based hybrid forecasting system. *ICAPR 2001, Lecture Notes in Computer Science*, 2013:72–82.

[92] H. M. T and M. M. B. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5:989–993, 1994.

[93] K. H. Talukder and K. Harada. Haar wavelet based approach for image compression and quality assessment of compressed image. *IAENG International Journal of Applied Mathematics*, 36, 2007.

[94] H. Tan. Neural network model for stock forecasting. Master's Thesis (1995), Texas Tech University.

[95] H. Tao. A wavelet neural network model for forecasting exchange rate integrated with genetic algorithm. *IJCSNS International Journal of Computer Science and Network Security*, 6:60–63, 2006.

[96] W. H. L. S. H. M. F. Tenorio. A neural network architecture for prediction. *Technical Report of the School of Electrical Engineering, Purdue University*, December, 1992.

[97] P. Tenti. Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence*, 10:567–581, 1996.

[98] P. A. C. M. G. A. J. G. C. V M Rivas, J J Merelo. Evolving rbf neural networks for time-series forecasting with evrbf. *Information Sciences*, 165:207–220, 2004.

[99] Y. N. Wei Huang, Kin Keung Lai and S. Wang. Forecasting foreign exchange rates with artificial neural networks: A review. *International Journal of Information Technology & Decision Making*, 3:145–165, 2004.

[100] B. Wu. Model-free forecasting for nonlinear time series (with application to exchange rates). *Computational Statistics & Data Analysis*, 19:433–459, 1995.

[101] C.-L. A. Xiang Li and R. Gray. An intelligent business forecaster for strategic business planning. *Journal of Forecasting*, 18:181–204, 1999.

[102] Y. R. Xin Wang and X. Shan. Wdrls: A wavelet-based on-line predictor for network traffic. *Proceedings of the IEEE GlobalCom 2003*, 7:4034–4038, 2003.

[103] J. D. Y Chen, B Yang and A. Abraham. Time-series forecasting using flexible neural tree model. *Information Science*, 174:219–235, 2005.

[104] H. Yamada. Do stock prices contain predictive information on business turning points? a wavelet analysis. *Applied Financial Economics Letters*, 1:19–23, 2005.

[105] B. Y. Yuehui Chen and A. Abraham. Optimal design of hierarchical wavelet networks for time-series forecasting. *In Proceedings of ESANN'2006*, pages 155–160, 2006.

[106] B. Y. Yuehui Chen and J. Dong. Nonlinear system modelling via optimal design of neural trees. *International Journal of Neural Systems*, 2:1–13, 2004.

[107] B. Y. Yuehui Chen and J. Dong. Time-series prediction using a local linear wavelet neural network. *Neurocomputing*, 69:449–465, 2006.

[108] L. P. Yuehui Chen and A. Abraham. Exchange rate forecasting using flexible neural trees. *Lecture Notes on Computer Science*, 3973:518–523, 2006.

[109] C. A. Zapart. On entropy, financial markets and minority games. *Physica A: Statistical Mechanics and its Applications*, 388:1157–1172, 2009.

[110] B.-L. Zhang and R. Coggins. Multiresolution forecasting for futures trading using wavelet decompositions. *IEEE Transactions on Neural Networks*, 12(4):765–775, 2001.

[111] B.-L. Zhang and Z.-Y. Dong. An adaptive neural-wavelet model for short term load forecasting. *Electric Power Systems Research*, 59:121–129, 2001.

[112] G. P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.

[113] G. Zumbach. Volatility processes and volatility forecast with long memory. *Quantitative Finance*, 4:70–86, 2004.